

This item is the archived peer-reviewed author-version of:

Assistive pointing device based on a head-mounted camera

Reference:

Walsh Edwin, Daems Walter, Steckel Jan.- Assistive pointing device based on a head-mounted camera
IEEE transactions on human-machine systems - ISSN 2168-2291 - 99(2017), p. 1-8
Full text (Publishers DOI): <http://dx.doi.org/doi:10.1109/THMS.2017.2649884>

Assistive Pointing Device Based on a Head-mounted Camera

Edwin Walsh, Walter Daems, Jan Steckel
 University of Antwerp
 FTI - Cosys / CZT
 Paardenmarkt 92 , B-2000 Antwerpen
 Belgium
 Edwin.Walsh@UAntwerpen.be

Abstract—This paper introduces, and validates the performance of, an alternative input device for people with limited hand/arm movement and control. A summary is provided of the current state of the art in alternative input devices. Based on this, a low cost solution is proposed that allows for (1) low latency, high operating speed and accuracy, (2) use from different viewing angles without recalibration, and (3) the ability to seamlessly control multiple devices. A prototype of this system was built and tested to determine the accuracy of the system (using a pan-tilt system), and to analyze the performance of the system compared to the state of the art (through user tests, based on the ISO 9241-411 test). The proposed system allows for great accuracy ($\sigma_X = 0.28 px, \sigma_Y = 0.29 px, \sigma_{XY} = 0.02 px$), a decent performance compared to the state of the art (Throughput = 1.52 bits/s, Error Rate = 31%), and good results in the ISO 9241-411 independent rating scale. The results from the performed experiments show the proposed system leads great promise in real-world applications. A low cost, head-mounted camera could be used as an alternative human interface device for people with limited hand/arm movement and control, allowing them to participate in the ongoing trend of computing devices gaining importance in our everyday activities.

Index Terms—assistive technology, optical sensor, head-pose tracking, pointing device, camera

I. INTRODUCTION

DUE to computing devices gaining importance in our everyday activities [1], an increased need arises to improve on human computer interaction (HCI) for people with a disability (that limits their ability to control computing devices), allowing them to participate in this growing trend [2], [3]. Typical HCI consists of soft controls (point-and-click interactions on graphical user interfaces) manipulated by physical controls (e.g. a mouse or keyboard). We define point-and-click operations as object selection and object interaction. Object selection is the process of selecting where on the screen an action should take place, e.g. by positioning the screen cursor or pointing a finger on a touch-screen. Object interaction is the process of performing an action on the selected object, e.g. clicking, scrolling, zooming, etc. While in mainstream applications object selection and object interaction are usually incorporated in the same device, it is the authors' view that due to the diverse nature of disabilities, methods for all input types should be considered separately. Hence, this paper will only focus on alternative methods for object selection. Also due to the diverse nature of disabilities, this

paper will be limited to a focus on a target audience of people with limited hand/arm movement or control.

There are many ways to characterize HCI [4]. One of these is pictured in Fig. 1. The user monitors the state of the computing device through sensory organs (vision, hearing, touch,...) and displays (visual, auditory, tactile,...), and controls the state through responders (limbs, voice, eyes, ...) and controls (mouse, touchpad, ...).

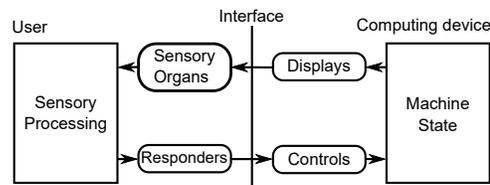


Fig. 1. Human-computer interface model based on [5].

The current main alternatives to traditional object selection methods for our target audience can be classified based on the order of controls [6], i.e. first-order control (velocity-control (1)) and zero-order control (position-control). Position-control can be subdivided into relative position control (2) (where displacement is measured) and absolute position control (3).

(1) *Velocity control*: With velocity control devices the responder-control interaction controls the velocity of the screen cursor [4]. Velocity control is often associated with switches or joysticks. Q Switch or joystick-oriented devices come in many forms including, but not limited to, traditional joysticks, tongue driven switches [7], accelerometer driven joysticks [8], or even through brain-computer interfaces [9]. The biggest issue with switches is that you can only move the screen cursor in one direction and with a fixed acceleration pattern. This pattern requires slow initial speed to allow small movement, leading to reduced operating speed of the system [10]. The problem with using joystick-oriented devices is that, although better operating speeds can be obtained than using switches [10], they are still perceived as slow, and can also be very intrusive and stigmatizing.

(2) *Relative position control*: With a relative position control device the sensed property is displacement (the amount of movement along a certain axis) [4]. Each sample is reported relative to the last sample. Relative position control is often associated with the traditional mouse or touchpads. For our

target group, a relative position control device exists using a gyroscopic sensor (e.g. QUHA), or with a (single marker) head-tracking system (e.g. HeadMouse Nano). The gyroscopic sensors work by sensing changes in orientation, the single marker head-tracking systems work by tracking a marker on the head using an optical or electromagnetic sensor system [11]. The main issue with these systems is the frequent need for clutching, and the difficulty of clutching. Clutching is a break in control-to-display mapping when the input device can be moved independently of the cursor (e.g. lifting the finger to reposition it on a trackpad) [12]. To perform clutching with the gyroscopic sensor or single marker head-trackers, an excessive corrective motion must be made by moving the head in the opposite direction of the desired movement. The need for this corrective motion is often seen as the major disadvantage for head-tracking systems using a single marker. An additional issue is that clutching becomes increasingly difficult if a dual-screen setup is used.

(3) *Absolute position control*: With an absolute position control device the position is sensed as an absolute coordinate at the point of contact. This point of contact can be a physical point of contact, e.g. the point where your finger touches a touch-screen, or virtual, e.g. the location on a screen where you are looking at (gaze tracking) or the location on a screen where your head is pointed at (head pose tracking). The previous examples use the two main responder types (eyes and head) available for our target group. We will first discuss head pose tracking, then gaze tracking.

Head pose tracking is the process of calculating the exact position of where the user's head is pointed at [13]. Head pose tracking has received much attention in the past decade, for its wide range of applications, such as assistive technology [14], [15], teleconferencing [16] and gaming [17]. Many researchers have focused on computer-vision-based head pose detection [18], but all these methods require powerful computational hardware, and they either report low angular accuracy, or make no statements about the accuracy. Simon Meers et al. [19] propose a head-tracking solution where three IR markers are placed on the user's head, combined with a webcam equipped with a visible light blocking filter. The advantage of this system is that it can be produced at a relatively low cost, and they report an angular accuracy of approximately 0.2° at 1 meter distance.

Gaze tracking is the process of calculating the exact position of what the user is looking at [20]. To calculate this, the exact position of the head should be known. This can be done by calibrating for a fixed head position, or by tracking the head position and rotation. Independently from the previously described methods for determining the position of the head, for gaze tracking, the most accurate methods have been video-based ones that determine the orientation of the eye by either using remote or head-mounted cameras [21]. The advantage of using remote cameras is that there is no requirement for wearability. The disadvantage is that one has to keep one's head in a fixed position, or that one has to use multiple cameras to track the movement and rotation of the head. There are several commercial eye gaze systems available, but all are still expensive (or require expensive software) and require the

user to be close to the camera. Although a lot of research has been done in search of a usable, low cost solution using gaze tracking [18], they all have their limitations, e.g. low accuracy, highly stigmatizing, CPU intensive. Furthermore, in spite of the high amount of research done, there are some disadvantages to using the eyes as an input device [22].

Firstly, due to the eye exhibiting a positional tolerance, it is not a highly accurate pointing device [23]. The fovea of the eye, which gives clear vision, covers a visual angle of approximately 1° arc of the retina, hence when fixating a target the eye only needs to be within approximately 1° of the target position to clearly see the target. This gives an inaccuracy in measured gaze position.

Secondly, the eye is being employed both as a sensory input method, so that the user can see the computing device's interface, and as responder output (see Fig. 1), indicating the pointing intention of the user to the interface [22]. This convergence of interaction point and gaze point means that the pointing cursor cannot be parked (left at a position on the screen while the eye momentarily looks away) to view the results of a user command or feedback from the interface. Such convergence results in possible unwanted and potentially distracting and unproductive pointing movements at the feedback point on the computing device as the cursor follows the eye wherever it gazes [23], [24]. Another issue is the saccadic nature of the sensory eye movement system [25] that impairs drawing smooth lines when controlling a screen cursor using gaze tracking.

II. PROPOSED SYSTEM

Based on our review of the state of the art [26], and discussions with industry experts, we think that an object selection method should aim for the following requirements:

- low latency, high operating speed and pointing accuracy
- usable from different viewing angles without recalibration
- able to control multiple devices
- low cost

In analogy with the system proposed by Meers et al. [19], we work with a cheap camera modified to block visible light but sensitive to IR light, using IR LEDs as markers. By using a filter to block visible light an image can be obtained which is entirely black with the IR LEDs appearing as bright intensity spots. The main difference with the solution proposed by Simon Meers is that the control unit is now carried by the user and only markers must be placed on the devices which can be operated through the control unit. This allows an easy extension enabling the control of multiple computing devices or environmental control systems, e.g. to turn on/off a light switch. As mentioned in the introduction, the proposed system has a focus on object selection, not object interaction. To allow for object interaction using the prototype, a push-button switch was used to simulate mouse clicks.

The initial prototype uses a Raspberry Pi NoIr (RPI) camera with a cutout from a floppy disk as visible light filter [19], and a wide-angle lens to allow for a larger field of view. IR LED markers are placed around the screen and their position relative to the top-left point of the screen is described

using homogeneous coordinates ($P_{l,W}$). The RPI camera takes snapshots which are sent to a computer running Matlab where blob detection is performed [27], noise is removed, and the detected blob centroids are sorted ($P_{l,C}$).

There are many different approaches possible for the placement of the markers and the calculation of the estimated location the camera is pointed at. Keeping into account possible future sensor-fusion, we decided to use four non-collinear points, so that the location and rotation of the camera compared to the screen can be calculated. For performance reasons, however, we estimate a homography matrix to define the relation between the coplanar marker coordinates and the detected blob coordinates [28]. This relation can be defined as

$$P_{l,W} = H \cdot P_{l,C} = \begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{pmatrix} \cdot P_{l,C} \quad (1)$$

with $P_{l,W} = (u \ v \ 1)^T$ the marker coordinates in the world coordinate system, $P_{l,C} = (x \ y \ 1)^T$ the detected marker coordinates in the camera coordinate system, and H the homography matrix.

An estimation of the homography matrix can be found using the Direct Linear Transformation (DLT) algorithm [28], by solving the following equation, for example by using a least squares method, for at least four points.

$$A_i h = 0 \quad (2)$$

with

$$A_i = \begin{pmatrix} -x_i & -y_i & -1 & 0 & 0 & 0 & u_i x_i & u_i y_i & u_i \\ 0 & 0 & 0 & -x_i & -y_i & -1 & v_i x_i & v_i y_i & v_i \end{pmatrix}$$

$$h = (h_1 \ h_2 \ h_3 \ h_4 \ h_5 \ h_6 \ h_7 \ h_8 \ h_9)^T$$

However, to ensure stability of the algorithm [28], both the matrices $P_{l,W}$ and $P_{l,C}$ should be normalized by computing the respective similarity transformations $T_{l,W}$ and $T_{l,C}$. These consist of a translation and scaling, that take the points from respectively $P_{l,W}$ and $P_{l,C}$ (after removing the homogeneous coordinate) to a new set of points $\tilde{P}_{l,W} = (\tilde{u} \ \tilde{v})^T$ and $\tilde{P}_{l,C} = (x \ y)^T$ such that the centroids of the points $\tilde{P}_{l,W}$ and $\tilde{P}_{l,C}$ are the coordinate origin $(0,0)^T$, and their average distance from the origin is $\sqrt{2}$. The DLT algorithm then consist of generating a $2n \times 9$ matrix \tilde{A} using A_i from formula (2) with the points from $\tilde{P}_{l,W}$ and $\tilde{P}_{l,C}$. After obtaining the SVD of \tilde{A} as $\tilde{A} = U D V^T$, where \tilde{h} can be found as the last column of V , the matrix \tilde{H} can be determined from \tilde{h} by reordering it into a square matrix as in (1). The homography matrix H can then be found as

$$H = T_{l,W}^{-1} \cdot \tilde{H} \cdot T_{l,C} \quad (3)$$

The intersection of the central axis of the camera with the plane defined by the markers ($P_{int,W}$) is calculated by transforming the central coordinates of the camera ($P_{c,C}$) using the estimated homography matrix.

$$P_{int,W} = H \cdot P_{c,C} \quad (4)$$

The X and Y coordinates of the intersection define the distance (in mm) to the top-left corner of the screen. These coordinates

are transformed into screen coordinates based on pixel size.

III. EXPERIMENTAL RESULTS

To validate the proposed system two different types of experiments were performed. First, to determine the accuracy of the system, the camera was placed on a PTU-D46 Pan-Tilt system (see Fig. 2.a) at four different locations around the screen, as pictured in Fig. 2.b. After these tests were performed, the system was validated by mounting the camera on a pair of glasses (see Fig. 3.a) and performing a multidirectional tapping task (see Fig. 3.b, based on the ISO 9241-411:2012(E) norm [29], using software provided by [30]). The results from this test showed the need for a filter to reduce the effect of unintentional head-movements. Finally, an improved version of the proposed system was tested by 18 healthy, untrained test participants. The test participants were asked to perform the multidirectional tapping task with the proposed system, and on devices from each of the defined input device categories, i.e. velocity control, relative position control, and absolute position control, to make a comparison with the current alternatives.

A. Accuracy of the system

In order to be able to point the camera at arbitrary points on the screen, we determined the settings of the pan/tilt system to direct the camera at 1600 coordinates evenly distributed over the screen. The error resulting from inaccuracies in the pan/tilt test setup was found to be negligible compared to the error resulting from the proposed implementation using the IR sensitive camera and LEDs. For each of these 1600 coordinates the screen cursor position was calculated 10 times based on the implementation using the IR sensitive camera and IR LEDs. Then the camera was directed to a randomly selected single coordinate on the screen and 1000 measurements were performed to define the standard deviation of the calculated screen cursor position.

Fig. 4.a shows the result of pointing the camera at 1600 target coordinates equally distributed over the screen from four different locations (see Fig. 2.b), and calculating the screen cursor position based on the proposed algorithm for 10 measurements per target coordinate. These locations were arbitrarily defined to indicate the relatively large operation range of the proposed system. The top row shows the mean error of the calculated screen cursor position for each target coordinate, the bottom row shows the standard deviation of the calculated screen cursor position for each target coordinate. The size of the mean error is relatively constant for each camera location, and changes smoothly over the 1600 calculated cursor positions, so that this error can be seen as a systematic error. Users automatically correct for this bias due to seeing the screen cursor position. A high standard deviation, however, would cause erratic behavior of the screen cursor. Hence a more important result is the relatively constant, and very small standard deviation for each calculated screen cursor position. While it was expected to see symmetrical results for positions C and D, this is clearly not the case. This result can be explained due to inaccuracies caused by the test process, where the calibration of the pan-tilt system to point the camera

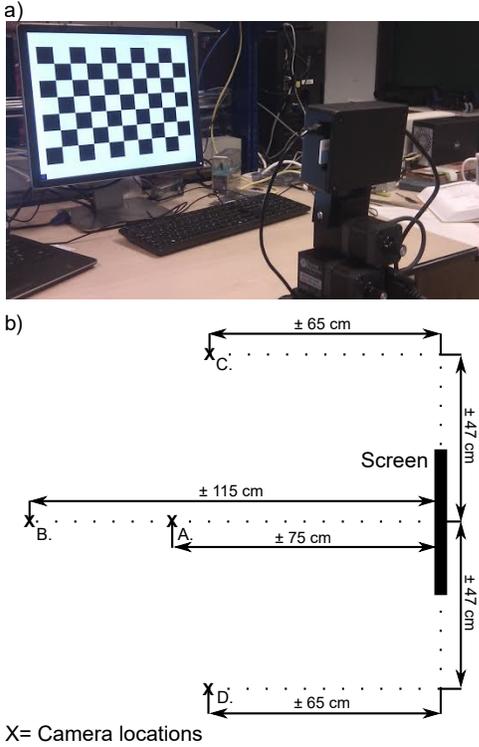


Fig. 2. a) Example of the PanTilt calibration setup, and b) camera locations for the system experiments.

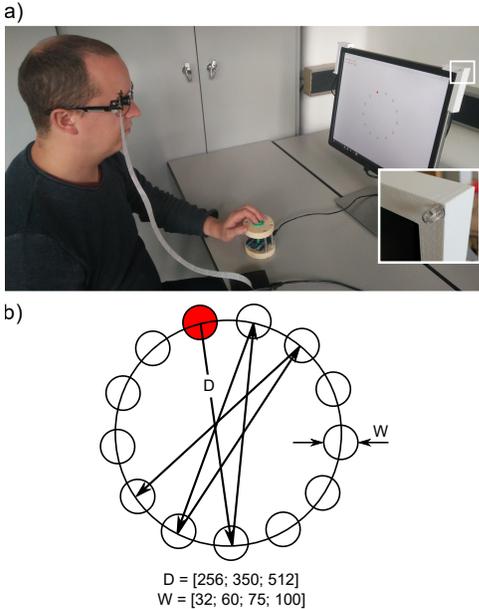


Fig. 3. a) An image of the sensors placed round the screen (with an added digital zoom), and the camera mounted on a pair of glasses as used during the trials, and b) a sketch of the ISO 9241-411 throughput testing scheme with a list of the used settings for the distance between the targets (D) and the size of the targets (W) in pixels.

at the 1600 coordinates must be done without IR bandpass filter, and the actual test must be performed with IR bandpass filter. To place the IR bandpass filter, the camera lens must be removed. Replacing the camera lens in a slightly different position can lead to this bias error. However, as mentioned

earlier in this paragraph, this bias error can be seen as a systematic error, for which a user will automatically correct when seeing the screen cursor. Fig. 4.b shows the direction of the errors for the measurements from location C. For clarity, the number of points were reduced, and the error was scaled by a factor 1.5. The error direction changes smoothly with respect to the cursor position which is desirable for smooth mouse operation. Fig. 4.c shows a histogram of 1000 estimated cursor positions, pointing the camera at a single random coordinate on the screen. The histogram indicates a very low standard deviation of the cursor position in pixels ($\sigma_X = 0.28, \sigma_Y = 0.29, \sigma_{XY} = 0.02$).

B. User validation

Next, the ISO 9241-411 multi-directional tapping test was performed by a limited set of four test participants using the settings as described in Fig. 3.b.

The multidirectional tapping task applies Fitts' law to evaluate input devices. During this test, the user has to 'tap' targets, positioned on a circle. Moving from one target to the next-opposite target is called a trial. Sequentially selecting all of the targets on the circle is called a sequence. For each test person, several iterations of this sequence have to be performed with varying target widths (area of the target) and amplitudes (distance between consecutive targets). After each trial a throughput is calculated to allow for comparison between different trials. The throughput gives an indication of the rate of information transfer when a user is operating an input device to control a cursor on a display and is calculated as

$$\text{Throughput} = \frac{I_{De}}{t_m} \quad (5)$$

where I_{De} is the effective index of difficulty, and t_m is the average time to complete a single trial. The time to complete a single trial is calculated from the initiation of movement of the input device to target selection¹. The effective index of difficulty is defined as

$$I_{De} = \log_2 \frac{d + w_e}{w_e} \quad (6)$$

where d is the distance of movement to the target, and w_e is the effective target width, which is the width of distribution of selection coordinates made by a participant during a tapping test. It is calculated as

$$w_e = 4,133 \cdot s_x \quad (7)$$

where s_x is the standard deviation of the selection coordinates in the direction where the movement proceeds. For more information we refer to the ISO 9241-411 norm.

We measured the throughput of four test participants with our head-mouse sensor and using a standard PC mouse as reference. Initial results (see Fig. 5 for more details) show an average throughput value of 1.4 bits/s using the proposed sensor, compared to an average throughput value of 4.3 bits/s

¹Please note this is not the exact definition as described in the ISO 9241-411 norm, since the definition used in the norm incorrectly refers to the time of a single trial, instead of the average time of a single trial.

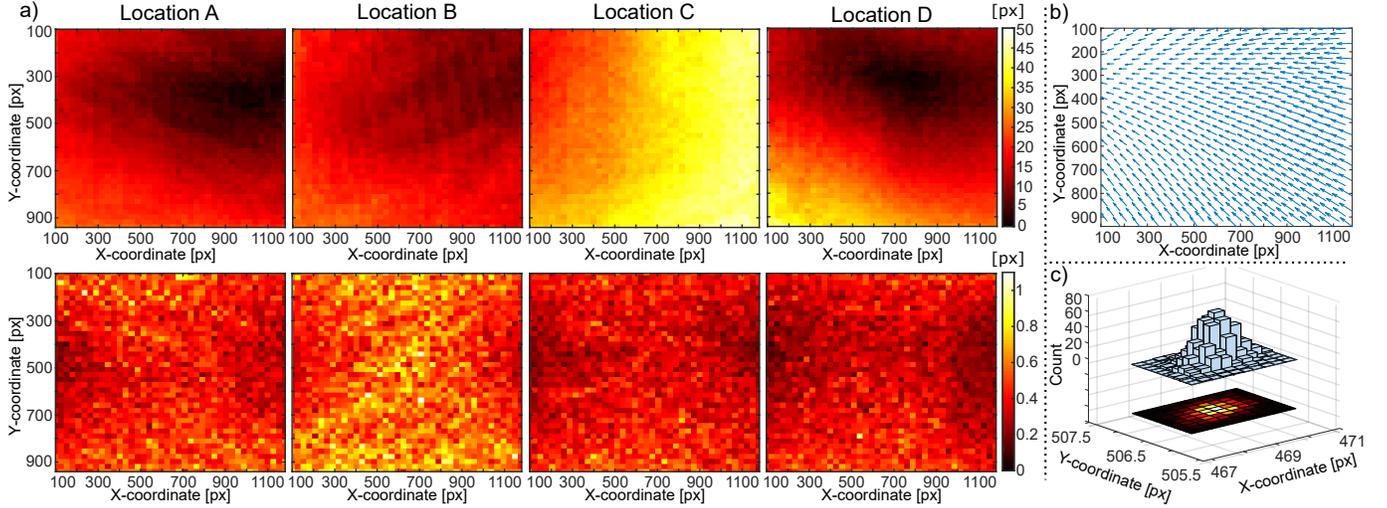


Fig. 4. a) Result obtained by pointing the camera from four different locations (see Fig. 2.b) at 1600 coordinates equally distributed over the screen, and calculating the screen cursor position based on the proposed algorithm for 10 measurements per coordinate. The top row shows the mean error of the calculated screen cursor position for each target coordinate, the bottom row shows the standard deviation of the calculated screen cursor position for each target coordinate. b) The direction of the errors for the measurements from location C (see Fig. 2). For clarity, the number of points were reduced, and the error was scaled by a factor of 1.5. c) Histogram of 1000 estimated cursor positions, pointing the camera at a single random coordinate on the screen.

using the standard mouse (to which the test participants were accustomed). These throughput values are in accordance with other state-of-the-art pointing devices (see ref. [30]–[33]).

Fig. 5 shows a 2D kernel density estimate of the normalized cursor position for a single user accumulated over all 12 sequences. The kernel density estimates, which show the relative occurrence of a certain cursor position, indicate a more linear path and higher accuracy during selection in case of the traditional mouse, as expected. While lower in accuracy, the head pose tracking sensor still yields adequate path linearity and accuracy. A more detailed description of these results can be found in [34].

While surveying the tests, we did notice, however, that test participants had problems with keeping the screen cursor completely still. We had a single user perform a test where the user was supposed to keep the screen cursor at a single location on the screen. Fig. 6.a and Fig. 6.c show the results of these tests measuring 1000 calculated cursor positions from a user trying to keep the cursor still at a single point. From Fig. 6.a we can conclude that the screen cursor was spread quite evenly over an area of 30x30 pixels. This can also be seen in Fig. 6.c, which shows the calculated positions compared to an (small) icon with the default size of 16x16 pixels. This shows that it would be very hard for a user to select small targets. While implementing a filter that could keep the screen cursor steady at a single point is quite a trivial task, implementing one that also maintains sufficient responsiveness is less trivial.

C. Filter

To deal with the high standard deviation and corresponding inability to select small targets we defined a custom Gaussian attractor. The Gaussian shape was selected for its smoothness. Since the custom Gaussian attractor displayed good experimental results, other filter types (e.g. non-linear signal processing techniques) were not investigated further.

The Gaussian attractor is used to calculate a filtered cursor position as

$$p'' = p + \alpha \cdot (p' - p) \quad (8)$$

with p the previous cursor location, p' the unfiltered estimated cursor position, and α a scaling factor ($\alpha \in [0, 1]$) that can be calculated as

$$\alpha = 1 - e^{-\frac{\|p' - p\|^2}{2\sigma^2}} \quad (9)$$

where

$$\sigma = \mu_T \cdot \rho \quad (10)$$

with μ_T the mean tremor experimentally defined for a single user and ρ an experimentally defined constant.

To test how stable a user could hold the screen cursor we had the same user try to hold the screen cursor on a single position, this time with the implemented filter activated. As can be seen in Fig. 6.b and Fig. 6.d the user has much finer control over the screen cursor position and is able to keep the screen cursor in an area of 16x16 pixels. To test the responsiveness of the system we had several users perform the multidirectional tapping task with the proposed system, and compared these to tests done with similar devices.

D. Comparison

We asked an additional 18 healthy, untrained test participants (2 female, 16 male, aging between 22 and 60 years old) to perform the multidirectional tapping task with devices from each of the defined input device categories, i.e. velocity control (QUHA [35] set up as a joystick), relative position control (HeadMouse Nano [36] and QUHA set up as a regular mouse), and absolute position control (Tobii eyeX [37]), to make a comparison between the improved proposed system (with filter) and the current alternatives. For each device, 5 multidirectional tapping test sequences were performed within each difficulty setting defined by the ISO 9241-411 norm as

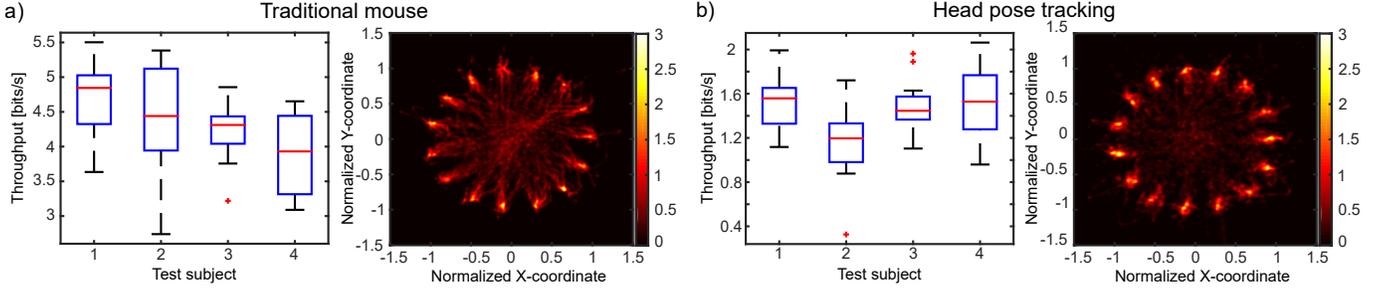


Fig. 5. Throughput and 2D kernel density estimate using a) a traditional mouse, and b) using the proposed system. The throughput graph is based on test results of four healthy untrained test participants performing 12 sequences of the ISO 9241-411 multi-directional tapping test. Each sequence consists of 13 trials with target distance and width as described in figure 3 executed in random order. The 2D kernel density estimate graph is based on the normalized cursor position for a single user accumulated over all 12 sequences. The kernel density estimates show the relative occurrence of a certain cursor position.

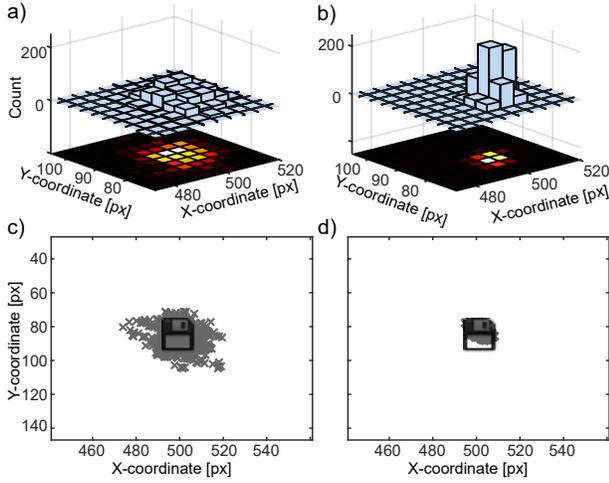


Fig. 6. 1000 calculated cursor positions from a user trying to keep the cursor still at a single point. a) Histogram of these cursor positions without using a filter. b) Histogram these cursor positions with a tremor filter activated. c) Scatter-plot of the cursor positions without using a filter compared to an icon of size 16x16 pixels. d) Scatter-plot of the cursor positions with a tremor filter activated compared to an icon of size 16x16 pixels.

seen in Table I. Leading to 20 test sequences executed in random order per person per device. Each participant tested each device in a random order. Each device was given a push-button switch to allow for simulation of mouse clicks. However, due to significantly lower throughput values of the velocity control system during preliminary testing, we left it out of the comparison. This due to the fact that it would take the test participants too long to perform the entire test series. The test participants were seated at roughly 1 m from a 24 inch screen with resolution 1280x1040.

TABLE I
MULTI-DIRECTIONAL TAPPING TEST WIDTHS AND AMPLITUDES

I_D	W	D_1	D_2	D_3	D_4	D_5
High ($I_D > 6$)	10	650	725	800	850	900
Medium ($4 < I_D \leq 6$)	20	450	550	650	750	850
Low ($3 < I_D \leq 4$)	60	425	525	625	725	825
Very low ($I_D \leq 3$)	100	500	550	600	650	700

Each user was given time to practice with each device until they felt they had sufficient control of the device. This practice time ranged from approximately 5 to 15 minutes. Each

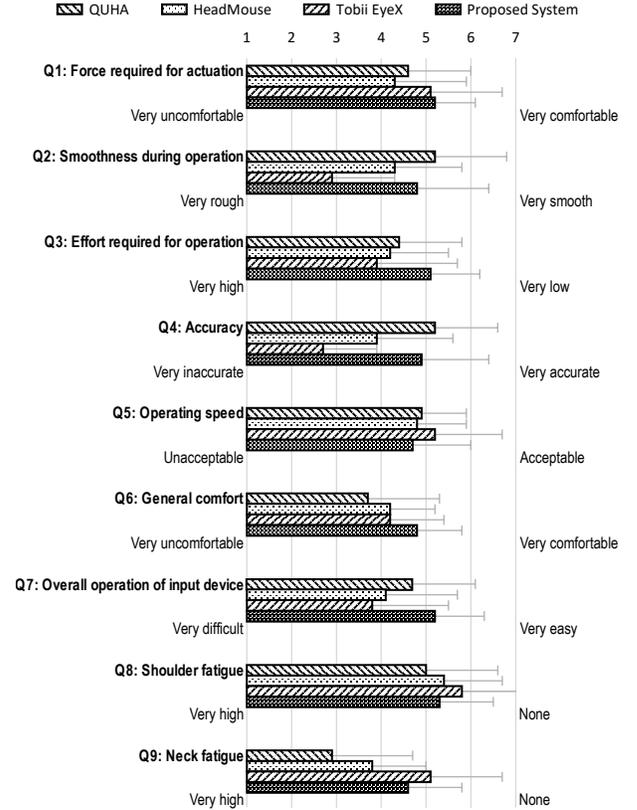


Fig. 7. Questionnaire results from questions presented to participants after completing all tests for each device. The horizontal bars show the mean value of the results, the error bars show the positive side of the standard deviation of the results.

device was set-up to comply as best as possible to the user's preferences. For the relative position control devices each user could setup his preferred gain factor (the mouse sensitivity), and for the absolute position control device, the device was calibrated for each individual user.

Additionally, we asked the participants to fill in a questionnaire based on the independent rating scale from the ISO 9241-411 norm [29], as depicted in Fig. 7, after performing all the tests for a single device. The questions were scored from 1 (the least favorable option) to 7 (the most favorable option).

The results of the multidirectional tapping test can be found in Fig. 8. The top row shows a box-plot of the mean throughput

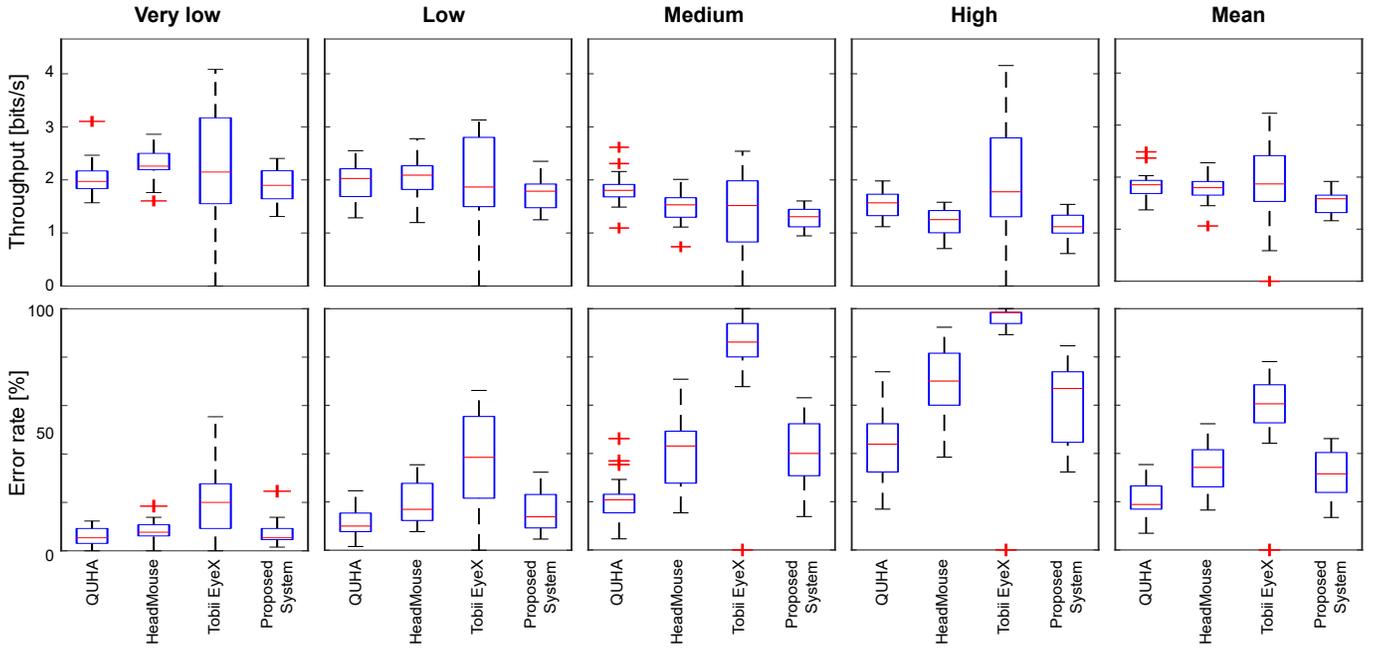


Fig. 8. Results of the multidirectional tapping test performed by 18 users with four head (or eye) based input devices. Each participant performed 5 tests for each of the selected difficulty settings on each device (see Table I). The top row shows a box-plot of the mean throughput per user for each device and each difficulty setting. The bottom row shows a box-plot of the error rate per user during these tests. The error rate is defined percentage wise as the number of times the participant missed the target compared to the total number of targets.

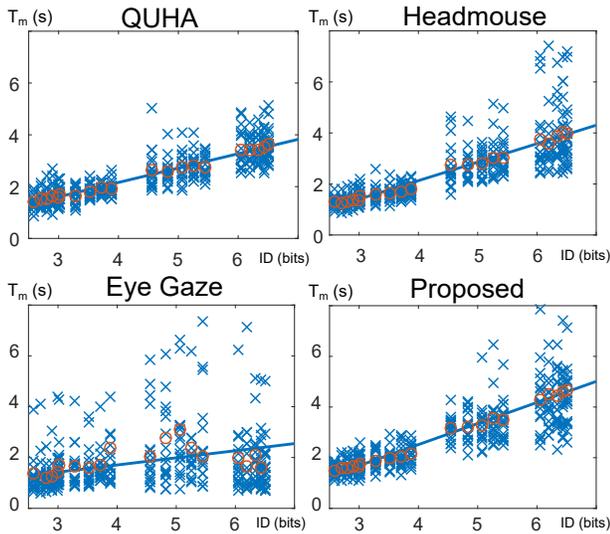


Fig. 9. Regression plot of the mean time to complete one trial per index of difficulty.

per user for each device and each difficulty setting. The bottom row shows a box-plot of the error rate per user during these tests. The error rate is defined as the number of times the participant missed the target compared to the total number of targets. Fig. 9 shows the regression plot of the mean time it took each participant to complete a trial for each ID setting. The results from the questionnaire can be found in the horizontal bar graph shown in Fig. 7.

The first thing of notice when looking at the throughput values is the high standard deviation for the Tobii EyeX system. While no irrefutable evidence can be given for the

reason of this high standard deviation, it is the authors view that this is caused by the way different users coped with the inaccuracies of the eye-gaze system. It was noted during the tests that some users had a larger focus on speed than on accuracy. This effect is shown in the extremely high error rates for the eye-gaze system, as seen in Fig. 8, and the large spread in mean movement time, as seen in Fig. 9. Due to this high standard deviation, the eye-gaze system will be left out of the statistical analysis, and we will continue with the head-controlled input devices.

Since the goal of this paper is not to make a full comparison of all tested systems, but to position the proposed system compared to commercially available systems, only a paired-samples t-test was conducted, with Bonferroni correction ($\alpha = 0.025$), to compare the mean throughput of the proposed system with the QUHA and HeadMouse Nano systems. There was a significant difference in the throughput value of the QUHA system ($\mu = 1.84, \sigma = 0.28$) and the proposed system ($\mu = 1.52, \sigma = 0.22$) conditions; $t(17)=6.24, p<0.001$. There also was a significant difference in the throughput value of the Headmouse Nano system ($\mu = 1.77, \sigma = 0.26$) and the proposed system ($\mu = 1.52, \sigma = 0.22$) conditions; $t(17)=4.64, p<0.001$. While these differences are statistically significant, it is the authors view that these differences are of little practical value. This is confirmed when analyzing the questions from the questionnaire.

A two-sided Wilcoxon signed-rank test was performed, with Bonferroni correction, for each of the questions to check if the difference between the results for the proposed system and the QUHA system, and the difference between the proposed system and the Headmouse Nano system, comes from a distribution with zero median. The results of these tests can

TABLE II
RESULTS FROM TWO-SIDED WILCOXON SIGNED-RANK TEST

	Proposed system							
	vs. QUHA				vs. Headmouse			
	W	p	h	Md	W	p	h	Md
Q1: Force required	74.5	0.176	0	0.0	47.0	0.055	0	0.0
Q2: Smoothness during operation	39.0	0.429	0	1.0	84.5	0.183	0	1.0
Q3: Effort required for operation	60.5	0.109	0	1.0	103.5	0.011	1	1.0
Q4: Accuracy	24.0	0.292	0	0.5	61.0	0.011	1	0.5
Q5: Operation speed	25.5	0.558	0	0.0	56.0	0.747	0	0.0
Q6: General comfort	99.0	0.026	0	1.0	98.0	0.027	0	1.0
Q7: Overall operation of input device	86.0	0.150	0	1.0	74.5	0.003	1	1.0
Q8: Shoulder fatigue	31.0	0.363	0	0.0	27.5	0.795	0	0.0
Q9: Neck fatigue	136.0	0.004	1	0.5	56.0	0.046	0	0.5

Legend: W is the test statistic for the Wilcoxon signed-rank test, p is the probability of observing the test statistic, h = 1 indicates a rejection of the null hypothesis, and h = 0 indicates a failure to reject the null hypothesis at the 2.5% significance level. Md is the difference from the median value of the questionnaire result from the compared systems.

be found in table II. It should be noted that compared to the QUHA system, the proposed systems scores statistically significantly better on question Q9, and no statistically significant difference is found for the other questions. Compared to the Headmouse Nano system, the proposed systems scores statistically significantly better on questions Q3, Q4 and Q7, and no statistically significant difference is found for the other questions.

It can be seen from the above analysis, that although the proposed system results in a statistically significantly smaller throughput value, the performance of the system is not perceived as lower. We would like to make special note of Q7, which indicates the overall performance of the device, where, although only statistically significant compared to the Headmouse system, the proposed system scores better than the other head-controlled systems.

If we would have been a little less conservative with the rejection criteria of the null hypothesis, and analyzed the questionnaire results with a significance level of 5%, we would come to the conclusion that there is a statistically significant difference between the proposed system and both the QUHA and Headmouse system for questions Q6 and Q9. These questions tell us something about the overall comfort. Again there is no irrefutable evidence to what is the cause for this result, but it is the authors view that this is caused due to the frequent need for clutching. While observing the test participants during the test, it was noted that many test participants limited the amount of clutching, which allowed for better test times, but which apparently led to a reasonable amount of strain on their neck muscles. This effect was not measured, but is a conclusion from the test participants' comments during the test and the questionnaire results.

IV. CONCLUSION

In this paper we have shown the need for a novel alternative input device for people with limited hand/arm movement based on a survey of the current state of the art targeted at our target audience. A new system was proposed based on IR markers, placed around a computing device screen, and a cheap IR sensitive camera mounted on a pair of glasses. It was shown that this system performs very well when placed on a pan-tilt system leading to very low standard deviations of the calculated cursor position in pixels ($\sigma_X = 0.28 px, \sigma_Y = 0.29 px, \sigma_{XY} = 0.02 px$). A Gaussian attractor filter was proposed to compensate for unintentional tremors during the participants' fine head movement. The proposed system was compared to current available commercial systems and was shown to achieve similar performance. It is expected that further improvements to the performance can be made, by improving the response time of the proposed system through sensor fusion with sensors with higher sample rates, like a gyroscopic sensor. Further analysis should also be done to investigate the impact of experience on the attainable throughput values for all compared systems.

The greatest benefit of the proposed system, compared to the commercially available systems however, is the benefit that the proposed system is easily extendable to allow control of multiple computing devices or even a home automation system. The used sensor (the camera), is perfectly capable of identifying at which marker the sensor is pointed at, using a simple modulation scheme, allowing for interaction with a great range of additional devices. It is the authors' view that the proposed system leads great promise in real-world applications.

ACKNOWLEDGMENT

This work is being supported by the Flemish Government (VAPH) as a part of the Flanders' care *Serenade* project. Special thanks go to Jabbla and Smartbox Assistive Technology, for their advise and support on available AT solutions.

REFERENCES

- [1] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media," *Business horizons*, vol. 53, no. 1, pp. 59–68, 2010.
- [2] K. M. Hjelle and K. Vik, "The ups and downs of social participation: experiences of wheelchair users in Norway," *Disability and rehabilitation*, vol. 33, no. 25–26, pp. 2479–2489, 2011.
- [3] K. Shinohara and J. O. Wobbrock, "In the shadow of misperception," in *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. New York, New York, USA: ACM Press, may 2011, p. 705.
- [4] I. S. MacKenzie, *Human-Computer Interaction: An Empirical Research Perspective*. Elsevier Science, 2012.
- [5] B. Kantowitz and R. Sorokin, *Human factors: understanding people-system relationships*. Wiley, 1983.
- [6] S. Zhai, "Human performance in six degree of freedom input controll," Ph.D. dissertation, University of Toronto, Department of Industrial Engineering, 1995.
- [7] J. Cheng, A. Okoso, K. Kunze, N. Henze, A. Schmidt, P. Lukowicz, and K. Kise, "On the tip of my tongue," in *Proceedings of the 5th Augmented Human International Conference on - AH '14*. New York, New York, USA: ACM Press, mar 2014, pp. 1–4.

- [8] I. S. MacKenzie and R. J. Teather, "FittsTilt," in *Proceedings of the 7th Nordic Conference on Human-Computer Interaction Making Sense Through Design - NordiCHI '12*. New York, New York, USA: ACM Press, oct 2012, p. 568.
- [9] R. Fazel-Rezai, B. Z. Allison, C. Guger, E. W. Sellers, S. C. Kleih, and A. Kübler, "P300 brain computer interface: current challenges and emerging trends." *Frontiers in neuroengineering*, vol. 5, p. 14, jan 2012.
- [10] S. K. Card, W. K. English, and B. J. Burr, "Evaluation of Mouse, Rate-Controlled Isometric Joystick, Step Keys, and Text Keys for Text Selection on a CRT," *Ergonomics*, vol. 21, no. 8, pp. 601–613, aug 1978.
- [11] V. Kindratenko, "A survey of electromagnetic position tracker calibration techniques," *Virtual Reality*, 2000.
- [12] M. Nancel, D. Vogel, and E. Lank, "Clutching is not (necessarily) the enemy," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, ser. CHI '15. New York, NY, USA: ACM, 2015, pp. 4199–4202. [Online]. Available: <http://doi.acm.org/10.1145/2702123.2702134>
- [13] E. Murphy-Chutorian and M. M. Trivedi, "Head pose estimation in computer vision: a survey." *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 4, pp. 607–26, apr 2009.
- [14] A. Sasou, "Acoustic Head Orientation Estimation Applied to Powered Wheelchair Control," in *Proceedings of the 2nd International Conference on Robotic Communication and Coordination*. IEEE, 2009, pp. 1–6.
- [15] Y. Song, Y. Luo, and J. Lin, "Detection of Movements of Head and Mouth to Provide Computer Access for Disabled," in *2011 International Conference on Technologies and Applications of Artificial Intelligence*. IEEE, nov 2011, pp. 223–226.
- [16] K. Otsuka, S. Araki, K. Ishizuka, M. Fujimoto, M. Heinrich, and J. Yamato, "A realtime multimodal system for analyzing group meetings by combining face pose tracking and speaker diarization," in *Proceedings of the 10th international conference on Multimodal interfaces - IMCI '08*. New York, New York, USA: ACM Press, oct 2008, p. 257.
- [17] D. A. Bowman and R. P. McMahan, "Virtual Reality: How Much Immersion Is Enough?" *Computer*, vol. 40, no. 7, pp. 36–43, jul 2007.
- [18] A. Al-Rahayfeh and M. Faezipour, "Eye Tracking and Head Movement Detection: A State-of-Art Survey," *IEEE Journal of Translational Engineering in Health and Medicine*, no. August, 2013.
- [19] S. Meers, K. Ward, and I. Piper, "Simple, Robust and Accurate Head-Pose Tracking Using a Single Camera," in *Mechatronics and Machine Vision in Practice*. Springer, 2008, pp. 111–122.
- [20] C. H. Morimoto and M. R. Mimica, "Eye gaze tracking techniques for interactive applications," *Computer Vision and Image Understanding*, vol. 98, no. 1, pp. 4–24, apr 2005.
- [21] V. Rantanen, T. Vanhala, O. Tuisku, P. Niemenlehto, J. Verho, V. Surakka, M. Juhola, and J. Lekkala, "A Wearable, Wireless Gaze Tracker with Integrated Selection Command Source for Human-Computer Interaction," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 15, no. 5, pp. 795–801, 2011.
- [22] R. Bates and H. Istance, "Why are eye mice unpopular? A detailed comparison of head and eye controlled assistive technology pointing devices," *Universal Access in the Information Society*, vol. 2, no. 3, pp. 280–290, oct 2003.
- [23] R. Jacob, "Eye tracking in advanced interface design," *Virtual environments and advanced interface design*, 1995.
- [24] B. Velichkovsky, A. Sprenger, and P. Unema, *Human-Computer Interaction INTERACT '97*, S. Howard, J. Hammond, and G. Lindgaard, Eds. Boston, MA: Springer US, jan 1997.
- [25] R. J. Leigh and D. S. Zee, *The neurology of eye movements*. Oxford University Press New York, 1999, vol. 90.
- [26] E. P. Walsh, J. Steckel, and W. Daems, "Human-computer interface using a head mounted camera and IR markers," in *Proceedings of the 13th AAATE Conference: Attracting new areas and building bridges*, 2015, pp. 593 – 596.
- [27] *MATLAB version 8.6.0.267246 (R2015b)*, The Mathworks, Inc., Natick, Massachusetts, 2015.
- [28] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.
- [29] "ISO 9241-411:2012(E) Ergonomics of human-system interaction – Evaluation methods for the design of physical input devices," International Organization for Standardization, Geneva, CH, Standard, Mar. 2012.
- [30] I. S. MacKenzie, "Evaluating eye tracking systems for computer input," *Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies: Advances in Assistive Technologies*, p. 205, 2011.
- [31] I. S. MacKenzie and R. R. J. Teather, "FittsTilt: the application of Fitts' law to tilt-based interaction," in *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*. ACM, 2012, pp. 568–577.
- [32] A. Pino, E. Tzemis, N. Ioannou, and G. Kouroupetroglou, *Human-Computer Interaction. Interaction Modalities and Techniques*, ser. Lecture Notes in Computer Science, M. Kurosu, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, jul 2013, vol. 8007.
- [33] S. P. Guness, F. Deravi, K. Sirlantzis, M. G. Pepper, and M. Sakel, "Evaluation of vision-based head-trackers for assistive devices." *Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, vol. 2012, pp. 4804–7, jan 2012.
- [34] E. Walsh, W. Daems, and J. Steckel, "An optical head-pose tracking sensor for pointing devices using IR-LED based markers and a low-cost camera," in *2015 IEEE SENSORS*. IEEE, nov 2015, pp. 1–4.
- [35] QUHA, "Quha zono," 2016, [Accessed: 2016-04-14]. [Online]. Available: <http://www.quha.com/products-2/zono/>
- [36] Origin Instruments, "Headmouse nano," 2016, [Accessed: 2016-04-14]. [Online]. Available: <http://www.orin.com/access/headmouse/>
- [37] Tobii AB, "Tobii eyex," 2016, [Accessed: 2016-04-14]. [Online]. Available: <http://www.tobii.com/xperience/products/>