It's in Your Eyes

Towards Context-Awareness and Mobile HCI Using Wearable EOG Goggles

It's in Your Eyes - Towards Context-Awareness and Mobile HCI Using Wearable EOG Goggles

Andreas Bulling
ETH Zurich
Wearable Computing Laboratory
bulling@ife.ee.ethz.ch

Daniel Roggen
ETH Zurich
Wearable Computing Laboratory

Gerhard Tröster
ETH Zurich
Wearable Computing Laboratory

ABSTRACT

In this work we describe the design, implementation and evaluation of a novel eye tracker for context-awareness and mobile HCI applications. In contrast to common systems using video cameras, this compact device relies on Electrooculography (EOG). It consists of goggles with dry electrodes integrated into the frame and a small pocket-worn component with a DSP for real-time EOG signal processing. The device is intended for wearable and standalone use: It can store data locally for long-term recordings or stream processed EOG signals to a remote device over Bluetooth. We describe how eye gestures can be efficiently recognised from EOG signals for HCI purposes. In an experiment conducted with 11 subjects playing a computer game we show that 8 eye gestures of varying complexity can be continuously recognised with equal performance to a state-of-the-art video-based system. Physical activity leads to artefacts in the EOG signal. We describe how these artefacts can be removed using an adaptive filtering scheme and characterise this approach on a 5-subject dataset. In addition to HCI, we discuss how this paves the way for EOG-based context-awareness, and eventually to the assessment of cognitive processes.

Author Keywords

Wearable Computing, Context-Awareness, Human-Computer Interaction (HCI), Electrooculography (EOG), Eye Gestures, Eye Tracking

ACM Classification Keywords

C.5.7 [Computer System Implementation]: Wearable Computers; H.5.2 [Information Interfaces and Representation]: User Interfaces—Interaction styles; I.5.4 [Pattern Recognition]: Applications—Signal Processing

INTRODUCTION

Activity recognition is a key mechanism to devise context-aware systems for mobile and ubiquitous computing. The recognition of physical activity in mobile situations, for example motion from body worn sensors, has been extensively studied. However, context-awareness encompasses more than mere physical activity. Aspects like user attention and intentionality remain mainly unexplored as these cannot be picked-up by sensors usually deployed in todays wearable and pervasive computing scenarios.

A rich source of information about the state of the user can be found in the movement of the eyes. This includes information related to the users' activities and their cognitive processes such as attention [10], saliency determination [7], visual memory [13] and perceptual learning [4]. As part of an ongoing project we seek to investigate up to which extent these aspects can be inferred from the analysis of eye motion to enable new kinds of context-aware applications. Explicit eye movements performed by the user can directly be used for HCI input. Mobile attentive user interfaces (MAUIs) may also infer user intention and activity or provide assistance to the user by analysing implicit eye movements.

In earlier work, we proposed Electrooculography (EOG) as a novel measurement technique for wearable eye tracking and the recognition of user activity and attention in mobile settings [3]. EOG, in contrast to well established vision-based eye tracking[1], is measured with body-worn sensors, and can be implemented as a wearable system. We envisioned wearable and unobtrusive EOG recordings to be implemented using electrodes integrated into glasses and the signals processed in real-time on a light-weight device worn on the body. In this paper we describe how this can be achieved, and further extend this work by demonstrating how "eye gestures" can be recognised from EOG for HCI purposes. The specific contributions of this work are (1) the design and implementation of a wearable EOG-based eye tracker for long-term recordings in daily life implemented as goggles, (2) an algorithm for continuous recognition of complex eye gestures from EOG signals, (3) a characterisation in a computer game where eye gestures are used for HCI and (4) the development and evaluation of a new method for artefact removal from EOG signals caused by walking.

RELATED WORK

Activity Recognition

Logan *et al.* studied activity recognition in an environment instrumented with a large number and variety of common sensors [11]. They found that among all activities, reading was one of the most difficult to detect. They concluded that in order to catch all types of physical activity in daily-life scenarios, novel sensors and algorithms need to be developed.

A growing number of researchers investigate movements of the eyes during daily activities. Important advances have

[1]With eye tracking we understand the recording of eye movements in contrast to gaze tracking which deals with gaze direction.

been made to understand how the human brain processes visual tasks [6], how vision contributes to the organisation of active tasks in everyday life [9] and how eye, head, and hand movements are coordinated temporally [16]. However, eye movements have not been used for activity or context recognition so far.

Eye-based Interaction

Eye gaze recorded using vision has long been investigated as a means to interact with a computer. Most HCI work has focused on direct manipulation of user interfaces (e.g. Zhai *et al.* [23]). Qvarfordt *et al.* explore an interactive human-computer dialogue system which uses eye gaze patterns to sense the users' interests [17]. Drewes *et al.* propose to use eye gestures consisting of several consecutive movements to implement new ways of human-computer interaction [5]. They show that these gestures are insensitive to accuracy problems, immune against calibration shift and do not exhibit the "Midas touch" problem (for details see [8]).

EOG-based Interfaces

Patmore *et al.* developed a system intended to provide a pointing device for people with physical disabilities [15]. Basic signal characteristics such as saccades, fixations and blinks have been used for controlling robots with the person remaining stationary [22, 21]. For mobile scenarios, similar characteristics were used to operate a wearable computer system for medical caregivers [14]. All of these studies show that EOG can be implemented as an easy to operate and reliable interface. While these systems use basic eye movements as an input, they do not make use of movement sequences to implement a more versatile input modality.

Eye Tracking Devices

The common method to track eye gaze in natural environments are systems based on video cameras. A number of commercial eye trackers are available of which some are targeted at mobile use, e.g. the *iView X HED* from SensoMotoric Instruments (SMI). Nevertheless, these systems still require bulky headgear and additional equipment to process the video streams. To our knowledge, at this stage no solution for video-based eye tracking exists that is unobtrusive enough to allow for long-term recordings while leaving the wearer unaffected during physical activity.
Several researchers investigate novel electrode configurations for wearable EOG recordings. A gaze detector which uses EOG electrode arrays mounted on ordinary headphones was proposed by Manabe *et al.* [12]. While this approach might be less obtrusive than electrodes sticked to the face, it turned out to raise other issues: Low signal-noise ratio (SNR) and poor separation of the movement components. Vehkaoja *et al.* presented a light-weight head cap with electrodes embroidered of silver coated thread [19]. A small device integrated into the cap allows for wireless data transmission. As yet it is still to be evaluated in operation.

WEARABLE ELECTROOCULOGRAPHY

Eye Movement Characteristics

The eyes are the origin of a steady electric potential field which can be described as a dipole with its positive pole at the cornea and its negative pole at the retina. The magnitude of this so-called corneoretinal potential (CRP) lies in the range of 0.4mV to 1.0mV. The CRP is the basis for a signal measured between two pairs of electrodes commonly placed above and below, and on the left and right side of the eye, the so-called Electrooculogram (EOG).

If the eyes move from the centre position towards the periphery, the retina approaches one of the electrodes while the cornea approaches the opposing one. This results in a change in the electric potential. Inversely, eye movements can be tracked by analysing these changes in the EOG signal. The electrode pairs capture the horizontal and the vertical component of eye motion. This requires good electrode placement, i.e. on the eyes' horizontal and vertical axes of motion, as otherwise increased crosstalk between both components occurs. Usually, the signal amplitudes resulting from horizontal eye movements are larger than those from vertical movements. Therefore, crosstalk affects the vertical component more severely. Signal crosstalk poses problems on robust detection of eye movement events and eye gaze tracking for which both components need to be analysed simultaneously.

In the human eye, only a small central region of the retina, the fovea, is sensitive enough for most visual tasks. This requires the eyes to move constantly as only small parts of a scene can be perceived with high resolution. Simultaneous movements of both eyes in the same direction are called *saccades*. Typical characteristics of saccadic movements are 400°/s for the maximum velocity, 20° for the amplitude and 80ms for the duration. *Fixations* are static states of the eyes during which gaze is held at a specific location. Humans typically alternate saccadic eye movements and fixations while perceiving their environment.

Design and System Architecture

The wearable eye tracking device was designed to fulfil the following requirements:

- **Wearable and light-weight** to achieve a convenient and unobtrusive implementation and minimise user distraction.
- **On-board data storage and low-power** to allow for autonomous long-term recordings in daily life.
- **Real-time capability** to be able to perform online signal processing directly on the device.
- **Acceleration and light sensors** to compensate for artefacts caused by physical activity and changes in ambient light [1].

Hardware

The hardware is made of two components (see Figure 1): *Goggles* with integrated electrodes and a signal processing unit (called *Pocket*) with a credit card size of 82x56mm. The *Pocket* can be worn on the body, e.g. in a cloth bag fixed to one of the upper arms (see Figure 7). The system weighs 208g and is powered by a 3.7V / 1500mAh Li-polymer battery attached to the *Pocket* which allows for more than 7 hours of mobile eye movement recording. Raw EOG signals can be recorded on two channels with a sampling rate of up to 250Hz and a resolution of 20 bits noise-free[2].

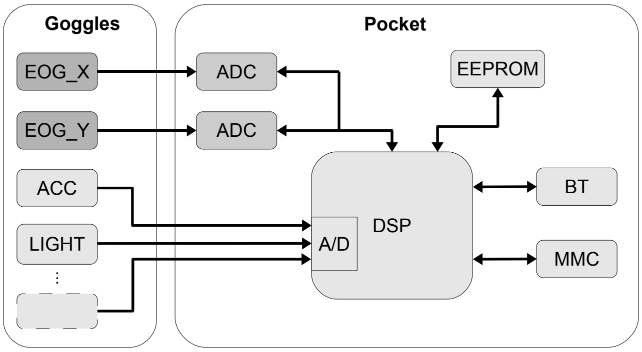

Figure 1. Two-part hardware architecture of the eye tracker with EOG amplification (EOG_X, EOG_Y), accelerometer (ACC), light sensor (LIGHT), DSP, analog-digital converters (ADC), EEPROM, Bluetooth module (BT) and MMC card holder.

The *Goggles* contain dry EOG electrodes and a small analogue amplification circuit board with a size of 42x15mm attached to the glasses frame. Four electrodes are arranged around the left eye and mounted on flat springs to achieve good skin contact. The EOG signal is composed of a small voltage superimposed by a large offset voltage relative to the ground electrode above the right eye. The offset is mostly caused by stray electrical signals on the leads and therefore referred to as common-mode interference. If an electric circuit is able to efficiently reject this interference it has a high *common-mode rejection ratio* (CMRR). To increase the CMRR, a *Driven Right Leg* (DRL) circuit [20] is implemented on the *Goggles*. Briefly, this circuit measures the common mode noise and feeds its negative back into the body to actively cancel the interference. Finally, an accelerometer and a light sensor are attached to the component with the latter pointing forward in line of incident light (see Figure 2).

The *Pocket* is the core signal processing unit of the system. It is based on a dsPIC micro-controller and contains two 24-bit analog-digital converters (ADC), a Bluetooth and a MMC module and an EEPROM. EOG signals coming from the ADCs are processed in real-time and can either be transmitted using Bluetooth or stored on the MMC. The EEPROM is used to store configuration data and parameters for the signal processing algorithms. Four LEDs and two buttons provide a simple interface which allows the user to access the basic functionality of the device.

Software
The dsPIC on the *Pocket* runs freeRTOS, an open-source real-time operating system devised for embedded systems. freeRTOS is configured to run in preemptive mode using predefined task priorities. The firmware is composed of three layers (see Figure 3). Among these layers, the *Hardware Abstraction Layer* (HAL) accesses the hardware. It provides a number of interfaces to the upper layers thus hiding all low-level hardware access. The *Device Layer* (DEL) uses the HAL to provide functionality for components external to the DSP such as the Bluetooth and the MMC module. The core functionality of the firmware is provided by 5 freeRTOS tasks which form the *Task Layer* (TAL). A separate *Library* (LIB) contains functionality which is shared by these tasks such as the CRC routines.

[2]The noise-free resolution of an ADC is the number of bits of resolution beyond which it is impossible to distinctly resolve individual outputs.

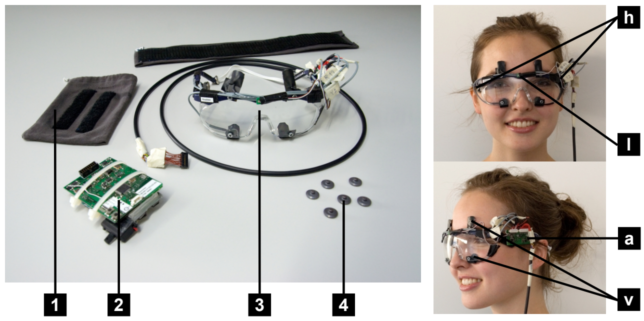

Figure 2. Components of the EOG-based wearable eye tracker: armlet with cloth bag (1), the *Pocket* (2), the *Goggles* (3) and dry electrodes (4). The pictures to the right show the *Goggles* worn by a person with the positions of the two horizontal (h) and vertical (v) electrodes, the light sensor (l) and the accelerometer (a).

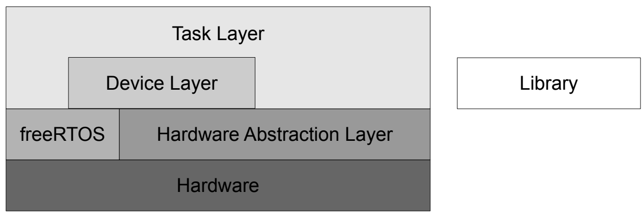

Figure 3. Three-tier software architecture of the eye tracker with layers used for hardware abstraction and task management by the operating system, access to external components and core functionality.

EOG SIGNAL PROCESSING
Eye gesture recognition is based on the detection of consecutive saccades which by their order and direction define the type of eye gesture. These saccades need to be detected in the continuous vertical and horizontal EOG signal streams. Blinks need to be removed because their characteristics are very similar to those of vertical eye movements and would affect gesture recognition. In this section we describe the processing steps required for the detection of blinks and saccades and the removal of blinks. We then describe the algorithms for eye gesture recognition and compensation of EOG signal artefacts induced by walking.

Blink Detection
We detect blinks with a template matching approach: First, a blink template is created using manually cutted equally-sized raw signal segments of 10 blinks from different persons, vertically shifted by their median and aligned at their peaks. To create the template, the mean at each sample point over all segments is calculated. Afterwards, blinks are detected by shifting this template over the vertical EOG sig-

nal component by following a sliding window approach. In each step, the Euclidean distance between the template and the signal segment of the current window is computed as a similarity metric. If the distance is below a defined threshold, i.e. the similarity between the template and the current segment is high, a blink event is recorded.

Saccade Detection

For saccade event detection we developed the so-called *Continuous Wavelet Transform - Saccade Detection* (CWT-SD) algorithm [2]: The CWT-SD first computes the continuous 1-D wavelet coefficients from the signal at scale 20 using Haar wavelets. A saccade event is detected for all samples where the absolute value of the coefficient vector exceeds a threshold. The direction and size of a saccade is given by its sign and amplitude.

Blink Removal

For blink removal, the streams of saccade and blink events are analysed in parallel. Blinks without simultaneous saccades are directly removed from the signal. In case of a simultaneous saccade, three cases need to be distinguished, namely *presaccadic blinks*, *postsaccadic blinks* and *intrasaccadic blinks*. To maintain the essential signal characteristics for the eye gesture recognition it is necessary to handle each of these in a specific way:

Presaccadic blinks are caused by blinks which share their last edge with a saccade. Presaccadic blinks are removed by replacing the blink interval with the signal value at the beginning of the blink.

Intersaccadic blinks usually occur during slow eye movements or fixation periods. This type of blink is removed by replacing its interval with a linear interpolation between the value at the beginning and the value at its end.

Postsaccadic blinks are blinks which immediately follow a saccade and thus share their first edge with it. For removal, the blink interval is replaced with the signal value at the end of the blink.

Eye Gesture Recognition

The idea of combining distinct relative movements to more complex eye gestures was introduced in [5] for a video-based eye tracker. We follow a similar approach for continuous recognition of eye gestures based on EOG: Our algorithm takes the streams of saccade events for the horizontal and the vertical signal component as its input. It distinguishes between eye movements in basic, diagonal and intermediate directions (see Figure 4, top right corner): Basic directions are left, right, up and down (L, R, U, D). Diagonal and intermediate eye movements are characterised by simultaneous saccades in both signal components but different angles (e.g. I, 9, J).

For each saccade event the algorithm uses a time window of 0.06s to check for a second event in the other component. If such an event is detected a diagonal eye movement has occurred. The algorithm then uses the saccades' amplitudes and signs to combine and map them to the appropriate diagonal direction. If a second event is not found within the given time window the initial saccade is directly mapped to the corresponding basic direction. This scheme assigns each eye movement a distinct event symbol, thus merging both signal components into one event string sequence (see Figure 4).

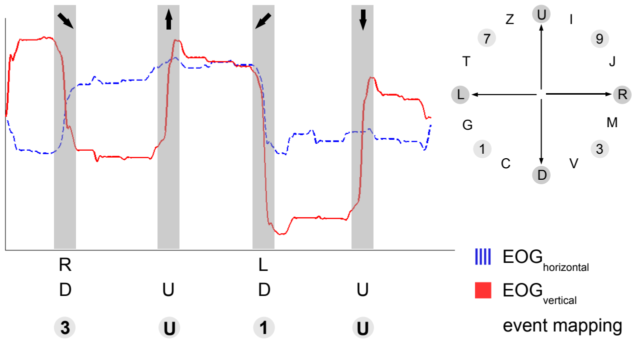

Figure 4. Eye movement event encoding from horizontal and vertical EOG signals for gesture 3U1U: Windows marked in grey with distinct eye movement events detected in the horizontal and vertical signal component and final mapping to basic (U) and diagonal (3, 1) movements. The top right corner shows the symbols representing the possible directions for eye movement encoding.

To recognise eye gestures which consist of several movements, the event string sequence is scanned for eye movement patterns following a string matching approach. To make the recognition more robust, the symbols representing intermediate directions are recognised as the nearest neighbouring symbol (e.g. I as U or 9). For matching, the current string sequence is continuously compared with string templates representing all possible gestures (see Table 1). For each template, the edit distance between the templates and the segment is calculated. If one of the templates exactly matches the current segment (i.e. the edit distance is zero), the corresponding eye gesture is recognised by the system.

Artefact Compensation

As EOG is measured with body-worn sensors, motion causes signal artefacts in the signals and affects eye movement detection. Walking is a common activity, e.g. on the way to work, during the day or in spare time at the weekend. Thus, walking serves as a good test bench for investigating artefacts induced by body motion. Analyses showed that artefacts in the EOG signals occur periodically according to the step frequency. A median filter with fixed window size fails to eliminate these artefacts for different persons and walking speeds. A parameter sweep on the window size using example data recorded from several subjects revealed that the optimal size is strongly related to the temporal step length. Therefore, we use an algorithm implementing an adaptive filter. The idea is to exploit the repetitive characteristic of walking and adapt the window size of the median filter to the step length as long as walking activity is detected (see Figure 5).

To detect walking, the algorithm first analyses the vertical

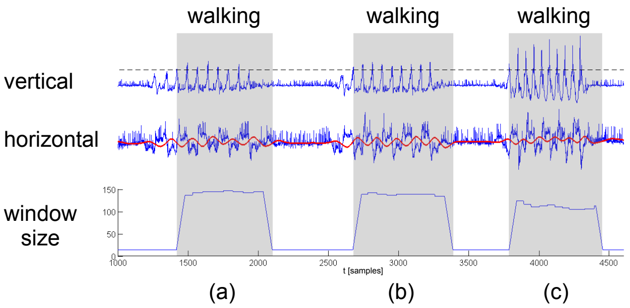

Figure 5. Adaptive filter for artefact compensation while walking slowly (a), moderate (b) and fast (c). Vertical acceleration signal and threshold (dashed line) for detecting walking activity. Horizontal acceleration signal and first derivative for calculating the step length. Resulting window size used by the median filter at the bottom.

Level 1	Level 2	Level 3	Level 4
R1R	DRUL	RDLU	RLRLRL
Level 5	Level 6	Level 7	Level 8
3U1U	DR7RD7	1397	DDR7L9

Table 1. Eye gestures of increasing complexity and their string representations used in the eight levels of the computer game (cf. Figure 4). The grey dot denotes the start and the arrows the order and direction of each eye movement.

axis of the goggle-mounted accelerometer. If the corresponding signal exceeds a defined threshold, the algorithm tries to detect steps by searching for zero-crossings of the first derivative of the low-pass-filtered acceleration data of the horizontal axis (see [18] for details). Walking is assumed as long as such steps are detected. In order to smooth out variations in walking style for different subjects, the step length is calculated on the basis of three consecutive step movements (e.g. right - left - right) separately for the left and the right leg. By calculating the length continuously for each step, the algorithm can adapt to different persons and walking speeds. For softer adaptation, only small increments are applied (see Figure 5). If walking activity is not detected anymore, the window size is incrementally set towards its default and static value.

EXPERIMENT I - EYE GESTURES FOR STATIONARY HCI

The aim of this experiment is to assess the feasibility of using the wearable electrooculographic system for stationary human-computer interaction. To investigate the use of explicit eye gestures, we developed a computer game consisting of eight different game levels. In each game level, subjects had to perform one defined eye gesture consisting of a changing number of consecutive eye movements. The gestures were selected and ordered to be of increasing complexity (see Table 1).

The first gesture was added to make it easier for the subject to familiarise with the game. Gesture 2 through 6 were inspired by [5] thus allowing to compare the results. The last two gestures contain shorter eye movements to assess if small-sized gestures can still be performed by the subjects and recognised by the system.

Setup

The experiments were conducted using the wearable eye tracker running at 100Hz sampling rate, a standard desktop computer and a 17" flat screen with a resolution of 1024x768 pixels. As reference points for the calibration procedure red dots were put at the corners and edges of the screen. The subjects were seated in front of the screen facing its centre (see Figure 6). In contrast to a previous study [5], no head stand was used, i.e. movements of the head and the upper body were allowed at any time during the experiments. However, we encouraged the subjects to sit upright with their eyes about 55cm to 65cm away from the screen. The expected movement order and their directions were shown as blue arrows with grey dots denoting the start and end point of a movement (see Figure 6 for an example).

Figure 6. Experimental setup consisting of a desktop computer running the game (1), the *Goggles* (2), a flat screen with red dots used for calibration (3) and the *Pocket* (4). The screenshots on the right show the sequence of eye movements and the generated event symbols for gesture 3U1U (from top to bottom). The red dot denotes the start of the gesture and the blue dot its end point. Blue arrows indicate the order and the direction of each expected eye movement and are masked out after having been successfully performed.

Experimental Procedure

In a first step, the classification thresholds were calibrated: The threshold for blink detection was determined by asking the subjects to blink 10 times and adjusting the threshold until the corresponding peaks in the EOG signal exceeded the threshold by about 30%. To calibrate for saccade detection, the subjects were asked to look in alternation at two of the red dots on opposite edges of the screen. To improve the calibration the same was repeated with an extra stopover at the centre of the screen. Afterwards, the recognition was verified by asking the subjects to focus on each red dot at the corners of the screen in clockwise order. The assistant then checked the stream of eye movement events for errors and initiated a re-calibration if necessary.

Gesture	Accuracy [%]										
	S1 (m*)	**S2 (f)**	**S3 (f)**	**S4 (m†)**	**S5 (m)**	**S6 (m*)**	**S7 (m)**	**S8 (m)**	**S9 (m)**	**S10 (m*)**	**S11 (m*)**
R1R	88	69	100	100	69	100	100	100	100	90	66
DRUL	100	71	100	100	86	100	100	90	100	100	90
RDLU	100	100	100	100	100	88	100	90	100	100	100
RLRLRL	100	100	95	93	100	100	92	82	100	89	88
3U1U	79	100	90	100	100	100	75	100	90	90	92
DR7RD7	90	95	78	71	90	93	88	75	96	80	100
1379	73	90	100	83	88	81	100	76	85	89	100
DDR7L9	95	91	93	71	89	77	100	73	76	76	100
Average	**91**	**89**	**95**	**90**	**90**	**92**	**94**	**86**	**93**	**89**	**92**

Table 2. Accuracy for the different gestures for each individual subject without test run. The accuracy gives the ratio of eye movements resulting in a correct gesture to the total number of eye movements performed. The table also shows the subjects' gender (f: female, m: male) and vision aid usually needed (*: glasses, †: lenses).

Once the calibration was successfully completed the experiment was started. The subjects performed three runs with all eight game levels being played in each run: The first was a test run to introduce the game and verify the blink and saccade thresholds for gesture recognition. No scores were recorded in this initial run. In two subsequent runs the subjects played all levels of the game again. In these two runs, the subjects were asked to concentrate on the game as performance measurements were taken to calculate the final game score.

In each game level, the corresponding eye gesture was to be repeatedly performed as fast as possible by the subject until the first successful try. To reach a high score, wrong eye movements, i.e. movements which were not part of the expected gesture, had to be minimised. For each correct movement, the corresponding arrow was masked out on the screen to reduce visual distraction. For the same reason, each correct and incorrect eye movement was indicated by a distinct sound. If a wrong eye movement was recognised, the level was restarted and a penalty was rewarded on the game score. Once a whole eye gesture was successfully completed, the next game level was started showing the next gesture. All wrong and correct eye movements as well as the time required to complete each gesture were recorded for each level. To trigger these measurements, the subject had to press a button once before performing each gesture. The total experiment time for each subject was about 25 minutes. At the end of the experiment, the subjects were asked on their experiences on the procedure in a questionnaire.

Results

We collected data from 11 subjects - 2 female and 9 male - between the ages of 24 and 64. (Originally there were 14 subjects, but 3 subjects had to be withdrawn due to poor signal quality resulting in calibration problems.) 4 subjects usually needed spectacles which they could not use during the experiment. For the subjects who completed the experiment, the average setup time was 7 minutes including putting on the glasses as well as setting up and calibrating the recording system.

The results for each individual subject only show a small range of different accuracies (see Table 2). The results were calculated solely using data from the second and the third run as the first one was only for testing. The accuracy was calculated as the ratio of eye movements resulting in a correct gesture to the total number of eye movements performed in the level. The highest result is 95% (subject 3) while the worst result was for subject 8, with an accuracy of 86%. It can be seen from the table that performance does not correlate to the gender of the subject. Also the datasets recorded from persons which usually need a vision aid do not show significant differences to the others.

The average performance over all subjects is given in Table 3 which shows the time, time ratio and the accuracy Acc to perform each of the eight gestures. T_T denotes the total time the subjects spent trying to complete each of the gestures while the success time T_S only measures the time spent on all successful attempts.

Gesture	$T_T[ms]$	$T_S[ms]$	T_S / T_T	$Acc[\%]$
R1R	3370	2890	0.858	85
DRUL	4130	3490	0.845	90
RDLU	3740	3600	0.963	93
RLRLRL	6680	5390	0.807	90
3U1U	4300	3880	0.902	89
DR7RD7	12960	5650	0.436	83
1379	6360	3720	0.585	84
DDR7L9	25400	5820	0.229	83

Table 3. Average performance for the different gestures over all subjects without test run. T_T is the total time spent to complete the gesture and T_S the success time spent only on successful attempts. The accuracy Acc is the ratio of eye movements resulting in a correct gesture to the total number of movements performed until success.

Table 4 shows the average time required to perform five gestures in comparison to a video-based system used in a previous study [5]. The raw times for EOG (T_R EOG) and video (T_R Video) show that the latter performs much better. However, the experimental setups differed in one important aspect: In the previous study all gestures were performed using a static interface. In this work the arrows indicating the direction were successively masked out as soon as the movement was recognised correctly. While this approach reduced visual distraction, it visually emphasised a characteristic of the system: Due to the signal processing involved, recognising one distinct eye movement took about half a second. We recognised that the subjects introduced a delay in waiting for the arrows to disappear. This obviously affected their overall response time. To take this into account, we assumed all movements had been performed without hiding the arrows. We estimated the response time T_R without delay as the average number of eye movements multiplied with the delay and subtracted from the average total time T_T.

Gesture	$T_R[ms]$		
	EOG	EOG w/o delay	Video
RDLU	3740	1630	1905
DRUL	4130	1520	1818
RLRLRL	6680	2860	3113
3U1U	4300	1940	2222
DR7RD7	12960	5890	3163

Table 4. Average response time T_R required to perform five different eye gestures over all subjects without initial test run in comparison to a video-based system. The third column gives estimated times correcting a difference in the experimental setups (see text for details).

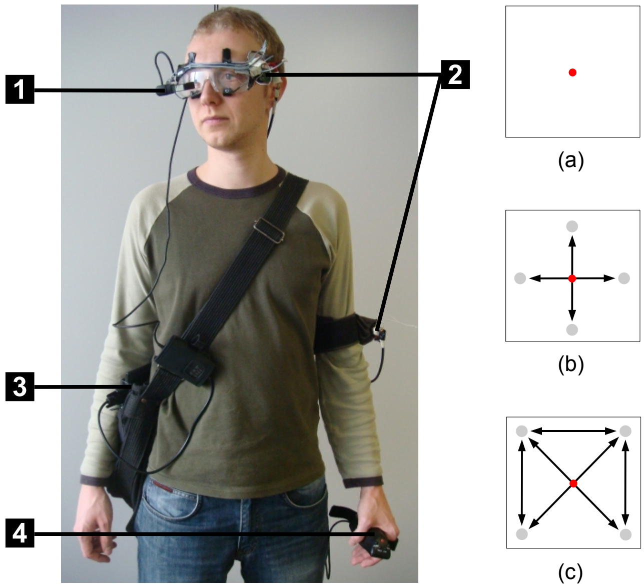

Figure 7. Experimental setup consisting of a head-up display (1), the wearable eye tracker (2), a laptop (3) and a Twiddler2 (4). The screenshots on the right show the different eye movements performed in the three runs: Fixations on the centre of the screen (a), simple movements in vertical and horizontal direction L'RLRLRL'U'DUDUDU' (b) and additional movements along the diagonals 7'R1U3ULD9D7R1' (c) (cf. Figure 4, quotation marks indicate movements of only half the distance). The red dots in the centre denote the start; arrows indicate the directions of the movements.

EXPERIMENT II - EYE MOVEMENTS FOR MOBILE HCI

In this experiment, we target a mobile scenario and investigate how artefacts induced by physical activity can be detected and compensated in EOG signals. The experimental scenario involved subjects to perform different eye movements on a head-up display (HUD) while standing and walking down a corridor. A custom software showed the expected movements with a defined order and timing.

Setup

The experiments were done using the wearable eye tracker running at 100Hz sampling rate, a standard laptop, a SV-6 head-up display from MicroOptical with a resolution of 640x480 pixels mounted to the *Goggles* frame and a wearable keyboard Twiddler2 (see Figure 7).

The laptop was used to run the experimental software. Similar to the first experiment, eye movements were indicated on the HUD as arrows with a red dot denoting the start and end point of each movement. During the experiments, the laptop was worn in a backpack in order not to constrain the subjects during walking. As the experimental assistant did not have control over the system, once the experiment was started, the Twiddler2 was needed to allow the subjects to control the software and start the different recordings.

Experimental Procedure

The subjects were not the same as for the first experiment, thus unfamiliar with the recording system. Therefore, they were first trained on the game using the laptop screen. Once the game was finished, the HUD was attached and the laptop was put in the backpack to start the experiment. The subjects performed three runs each consisting of different visual tasks while standing and walking down a corridor (see Table 5). A moving dot indicated the sequence and direction of the expected eye movements for each of these tasks. The subjects were asked to concentrate on their movements and fixate this dot permanently. The timing was exactly specified which resulted in one eye movement about every 3s.

The first run was carried out as a baseline case with fixations on the centre of the screen and large saccades without using the HUD. In two subsequent runs the subjects were asked to perform different sequences of eye movements on the HUD while standing and walking: The second run only contained simple movements in vertical and horizontal direction. The third run also included additional movements along the diagonals. Starting in the centre of the screen, these two sequences encode to L'RLRLRL'U'DUDUDU' and 7'R1U3ULD9D7R1' (cf. Figure 4, quotation marks indicate movements of only half the distance).

Results

We recorded 5 male subjects between the age of 21 and 27 totalling roughly 35 minutes of recording with walking activity accounting for about 22 minutes. As the mobile setting did not allow to record a ground truth, we decided to do a comparison to assess a relative performance measure. In a first step, the total number of detected saccades was calculated using the raw data of all subjects. This was done separately for run 2 and 3, for both standing and walking, for each subject and for the horizontal and the vertical signal component. The thresholds for the saccade detection algorithm were fixed to $T_{saccH} = 700, T_{saccV} = 2000$ for all subjects and runs. This analysis was then repeated twice with the same data: Once filtered by a median filter on a sliding window with a fixed size of 20 samples (0.2s) and once

Run	Visual task	Activity
1	1. large horizontal eye movements	standing
	2. large vertical eye movements	standing
	3. fixation	walking
	4. fixation	walking
2	1. simple eye movements	standing
	2. simple eye movements	walking
	3. simple eye movements	walking
3	1. complex eye movements	standing
	2. complex eye movements	walking
	3. complex eye movements	walking

Table 5. Experimental procedure consisting of three runs with performing different eye movements: Baseline measurement with large eye movements without using the head-up display and fixations, simple (run 2) and complex (run 3) eye movements each while standing and walking down a corridor.

after applying the adaptive filter. As the stationary results were used as a reference the fixed window size was selected to show good results for these recordings.

Figure 8 shows boxplots for the total number of detected saccades in the horizontal EOG signal component of run 3. Each box summarises the statistical properties of the data of the 5 subjects: The horizontal red lines in each box indicates the median and the upper and lower quartiles. The vertical dashed lines indicate the data range, points outside their ends are outliers. Boxes are plotted for the following cases: stationary and raw signal, stationary and fixed median filter, stationary and adaptive filter, walking and raw signal, walking and fixed median filter, walking and adaptive filter. The single solid horizontal line indicates the expected number of saccades defined by the experimental procedure.

What can be seen from the figure is that, in general, more artefacts are detected as saccades in the vertical EOG signal component. In the stationary case, both filters perform equally well but, compared to the expected number of saccades, improve the results only slightly. During walking, however, significant differences can be recognised: The raw recordings show about eight times more detected saccades than in the stationary case which renders eye movement detection impossible. While the median filter with a fixed window size fails in removing these artefacts, the adaptive filter still performs well, particularly for the horizontal EOG signal component.

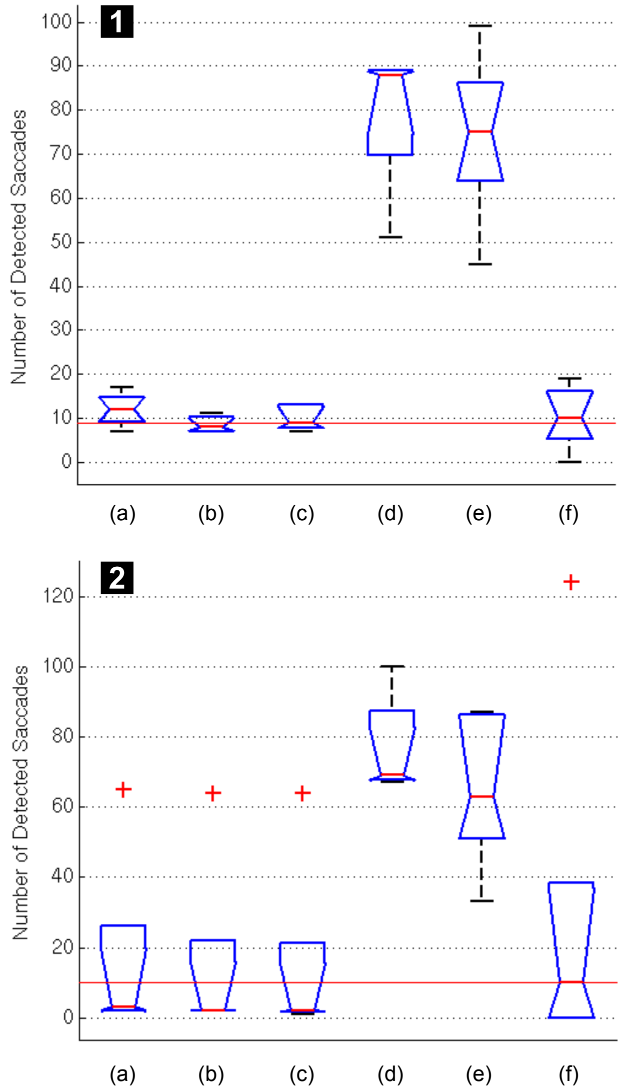

Figure 8. Boxplots for the total number of detected saccades in the horizontal (1) and vertical (2) EOG signal component of run 3 with fixed thresholds $T_{saccH} = 700, T_{saccV} = 2000$ over all subjects: stationary/raw (a), stationary/fixed median filter (b), stationary/adaptive filter (c), walking/raw (d), walking/fixed median filter (e), walking/adaptive filter (f). Horizontal red lines in each box indicate the lower quartile, median and upper quartile; dashed vertical lines show the data range; outliers are given as red crosses; the single solid horizontal line indicates the expected number of saccades.

DISCUSSION

On Eye Tracking Using EOG

In this work, a novel wearable eye tracker was described and evaluated. In contrast to common solutions using video, which require rather bulky equipment, this compact device is based on EOG. This enables a light-weight and unobtrusive integration into goggles which makes the system suited for mobile recordings in daily-life. As EOG requires much less computational power, this allows for low-power design and on-board storage which are crucial points for autonomous long-term recordings.

Our results show that EOG performs equally well to video if gaze tracking is not required. In the same way as video-based systems, EOG requires a calibration procedure. Once the thresholds are set, EOG is robust to varying distances between the person and the screen. In mobile settings with different screen sizes, the algorithm for eye gesture recognition using video introduced in [5] may require the grid size and timeout parameters to be continuously adapted. This is difficult to achieve automatically without knowing the screens' dimensions and the user's relative distance. With EOG, the thresholds only need to be adapted if the screen size is reduced considerably. By implementing a procedure which automatically detects if a recalibration is required, adaptation can be performed in the background without distracting the subject.

Wearability and comfort are important for long-term use. Nine subjects from our study reported that they felt uncomfortable due to rather high electrode pressure especially below the eye. In the questionnaire, however, they did not report of having felt physically constrained or distracted during the game by wearing the goggles. For 3 out of 14 subjects eye movements could not be detected. For the first subject, the goggles did not fit well which resulted in a lot of crosstalk in the vertical signal component. Suitable thresholds could not be found as the crosstalk was almost as strong as the vertical signal itself. For the second subject, probably due to dry skin, the vertical signal component was poor even though the goggles did fit well. The third subject declared afterwards having been very tired during the experiment. This could clearly be seen in the EOG signal by the presence of a lot of blinks and correctional saccades. These artefacts could not be removed completely and rendered eye gesture recognition impossible. We aim to solve these problems with the next prototype of the eye tracker which is currently under development. This includes revised goggles which can more easily be adapted to individual differences in head shape and size and which provide a mechanically improved electrode mounting.

On EOG-Based Eye Gestures

From the first experiment we found that EOG is a potentially robust input modality for HCI applications. EOG signals can be efficiently processed to recognise even complex gestures consisting of different consecutive eye movements. Table 4 shows that a set of gestures used to play a computer game can be recognised with equal performance to a video-based system. It has to be noted, however, that we estimated part of the results due to a different experimental setup. Still, we are confident this estimation is reasonable on average and that the eye tracker would still perform comparably after changing this setup.

The concept of playing a computer game using eye gestures was quickly understood by all subjects. We see in Table 2 that all of them were able to achieve a eye gesture accuracy of around 90% and often managed to perform the various gestures at the first try. Surprisingly, the accuracy for the easiest gesture in the first game level was similar to the last two (see Table 3). A possible explanation for this might be that the players were more inexperienced at the beginning and needed time to accustom to the game.

Although not shown here, from their performance we found that all subjects quickly learned how to use their eyes as a control input. However, using explicit eye gestures remained odd and 30% of the subjects reported of having had problems to stay concentrated during the experiment. They accounted this to the fact that controlling their eye movements consciously was a bit tiring. However, fatigue is an intrinsic problem not only for eye gestures but also for common input modalities such as speech or hand gestures. Eye gestures outperform these modalities if the hands can not be used (e.g. during driving, during a surgery or while working on the computer) or if speech input is not possible (e.g. for privacy reasons or in very silent or very noisy surroundings).

Six subjects usually needed vision aids which they could not use during the experiment. Surprisingly, Table 2 shows that these subjects performed equally well compared to those with normal sight. At least for the distance between the subject and the screen used in the experiment, the missing sight correction did not prevent to perform the different gestures successfully. However, it is clear that with a view to long-term use in ubiquitous settings with a variety of interfaces in different distances, goggles which still allow to use spectacles at the same time are desirable.

By analysing the overall performance we uncovered an interesting result: Gestures only consisting of large movements in the horizontal, vertical and diagonal directions worked well while those with smaller scale were more difficult to detect reliably. This is probably caused by the low amplitudes in the vertical EOG signal also recognised in the second experiment. This might be solved by simple means, e.g. by optimising the electrode placement and mounting or by designing HCI interfaces which rely on eye gestures performed mostly on the horizontal axis or with larger vertical eye movements. For mobile HCI with simultaneous physical activity, these aspects will become even more important.

On Artefact Compensation

From the experiments we found that for EOG recordings, particularly in mobile settings, efficient algorithms able to cope with signal artefacts caused by physical activity are required. Without compensation, artefacts may dominate the signal which renders eye movement detection impossible. Figure 8 shows that the proposed adaptive filter tuned to the walking pace can remarkably reduce the number of artefacts caused by walking activity. For long-term eye movement recordings with a wide range of different activities which constantly change during the day, more complex algorithms are clearly needed.

Conclusion

In this work we have shown the feasibility of building an autonomous eye tracker based on EOG. The device can be worn on the body which makes it particularly suited for long-term eye movement analysis in daily-life. A major benefit of EOG lies in the minimal amount of power and computation that is required for signal processing. By connecting several eye trackers, concurrent eye movement recordings for several people and distributed activity recognition may become possible. We have also shown that recognition of explicit eye gestures from EOG can be implemented as efficiently and robustly across different subjects as for video-based systems. EOG-based eye input allows for versatile human-computer interaction and may eventually provide new means of light-weight interaction in mobile settings by complementing current input modalities.

Our long-term objective is to investigate how much information eye motion can provide about the user's activity and context. In addition to further HCI refinements, we plan to investigate unconscious eye movements. Unconscious eye movements are the result of cognitive processes in the human brain. These processes are related to external aspects such as the user's activity or his environment, but also to internal aspects of visual perception, memory and learning. The analysis of eye motion thus may allow to access these underlying cognitive aspects. This would give important input for future context-aware systems.

ACKNOWLEDGEMENTS

We would like to thank Roberto Celli, Christian Meyer and Roman Suter for their work on building the device, implementing the signal processing algorithms and conducting the experiments.

REFERENCES

1. M. Brown, M. Marmor, Vaegan, E. Zrenner, M. Brigell, and M. Bach. ISCEV Standard for Clinical Electro-oculography (EOG), 2006. *Documenta Ophthalmologica*, 113(3):205–212, 2006.
2. A. Bulling, P. Herter, M. Wirz, and G. Tröster. Automatic Artefact Compensation in EOG Signals. In *Adj. Proc. of the 2nd European Conference on Smart Sensing and Context (EuroSSC 2007)*, pages 12–13, 2007.
3. A. Bulling, J. A. Ward, H.-W. Gellersen, and G. Tröster. Robust Recognition of Reading Activity in Transit Using Wearable Electrooculography. In *Proc. of the Sixth International Conference on Pervasive Computing (Pervasive 2008)*, pages 19–37, 2008.
4. M. M. Chun. Contextual cueing of visual attention. *Trends in Cognitive Sciences*, 4(5):170–178, 2000.
5. H. Drewes and A. Schmidt. Interacting with the Computer Using Gaze Gestures. In *Proc. of the 11th International Conference on Human-Computer Interaction (INTERACT 2007)*, pages 475–488, 2007.
6. M. Hayhoe and D. Ballard. Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9:188–194, 2005.
7. J. M. Henderson. Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, 7(11):498–504, 2003.
8. R. J. K. Jacob. What you look at is what you get: eye movement-based interaction techniques. In *Proc. of the SIGCHI conference on Human factors in computing systems (CHI 1990)*, pages 11–18, 1990.
9. M. F. Land. Eye movements and the control of actions in everyday life. *Progress in Retinal and Eye Research*, 25(3):296–324, 2006.
10. S. P. Liversedge and J. M. Findlay. Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, 4(1):6–14, 2000.
11. B. Logan, J. Healey, M. Philipose, E. Tapia, and S. Intille. A Long-Term Evaluation of Sensing Modalities for Activity Recognition. In *Proc. of the 9th International Conference on Ubiquitous Computing (UbiComp 2007)*, pages 483–500, 2007.
12. H. Manabe and M. Fukumoto. Full-time wearable headphone-type gaze detector. In *Ext. Abstr. of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2006)*, pages 1073–1078, 2006.
13. D. Melcher and E. Kowler. Visual scene memory and the guidance of saccadic eye movements. *Vision Research*, 41(25-26):3597–3611, 2001.
14. F. Mizuno, T. Hayasaka, K. Tsubota, S. Wada, and T. Yamaguchi. Development of hands-free operation interface for wearable computer-hyper hospital at home. In *Proc. of the 25th Annual International Conference of the Engineering in Medicine and Biology Society (EMBS 2003)*, pages 3740–3743, 2003.
15. D. W. Patmore and R. B. Knapp. Towards an EOG-based eye tracker for computer control. In *Proc. of the 3rd International ACM Conference on Assistive Technologies (Assets 1998)*, pages 197–203, 1998.
16. J. Pelz, M. Hayhoe, and R. Loeber. The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research*, V139(3):266–277, 2001.
17. P. Qvarfordt and S. Zhai. Conversing with the user based on eye-gaze patterns. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2005)*, pages 221–230, 2005.
18. G. Schindler, C. Metzger, and T. Starner. A Wearable Interface for Topological Mapping and Localization in Indoor Environments. In *Proc. of the 2nd International Workshop on Location- and Context-Awareness (LoCA 2006)*, pages 64–73, 2006.
19. A. T. Vehkaoja, J. A. Verho, M. M. Puurtinen, N. M. Nojd, J. O. Lekkala, and J. A. Hyttinen. Wireless Head Cap for EOG and Facial EMG Measurements. In *Proc. of the 27th Annual International Conference of the Engineering in Medicine and Biology Society (EMBS 2005)*, pages 5865–5868, 2005.
20. J. G. Webster. *Medical Instrumentation: Application and Design*. John Wiley and Sons, New York, 4th edition, 2007.
21. W. S. Wijesoma, Kang S. W., Ong C. W., A. P. Balasuriya, Koh T. S., and Kow K. S. EOG based control of mobile assistive platforms for the severely disabled. In *Proc. of the International Conference on Robotics and Biomimetics (ROBIO 2005)*, pages 490–494, 2005.
22. C. Yingxi and W. S. Newman. A human-robot interface based on electrooculography. In *Proc. of the International Conference on Robotics and Automation (ICRA 2004)*, pages 243–248, 2004.

23. S. Zhai, C. Morimoto, and S. Ihde. Manual and gaze input cascaded (MAGIC) pointing. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI 1999)*, pages 246–253, 1999.