Robust Recognition of Reading Activity in Transit Using Wearable Electrooculography

Andreas Bulling[1], Jamie A. Ward[2], Hans Gellersen[2] and Gerhard Tröster[1]

[1] ETH Zurich, Wearable Computing Laboratory
bulling@ife.ee.ethz.ch
[2] Lancaster University, Embedded Interactive Systems Group
j.a.ward@lancaster.ac.uk

Abstract. In this work we analyse the eye movements of people in transit in an everyday environment using a wearable electrooculographic (EOG) system. We compare three approaches for continuous recognition of reading activities: a string matching algorithm which exploits typical characteristics of reading signals, such as saccades and fixations; and two variants of Hidden Markov Models (HMMs) - mixed Gaussian and discrete. The recognition algorithms are evaluated in an experiment performed with eight subjects reading freely chosen text without pictures while sitting at a desk, standing, walking indoors and outdoors, and riding a tram. A total dataset of roughly 6 hours was collected with reading activity accounting for about half of the time. We were able to detect reading activities over all subjects with a top recognition rate of 80.2% (71.0% recall, 11.6% false positives) using string matching. We show that EOG is a potentially robust technique for reading recognition across a number of typical daily situations.

1 Introduction

Activity recognition has recently emerged as a key area of research in building context-aware interfaces for mobile and pervasive computing. The problem of recognising physical activity in mobile situations, for example using body worn sensors, has been investigated by several researchers [17,21]. However, recognition of activities based on more subtle cues, such as user attention and intention - a far more difficult problem - remains relatively unexplored.

A rich source of information on user activity is in the movement of the eyes. The paths that our eyes follow as we carry out specific activities also reveal much about the activities themselves. This is particularly true for activities with very specific eye movements, such as reading. Reading is a pervasive activity, e.g. on computer screens at work, advertisements and signs in public, and books read at home or while travelling. Thus information on a person's reading activities can be a useful indicator of his daily situation as well as a gauge of task engagement and attention. Attentive user interfaces could comprise the current level of user interruptability or provide assistance to people with reading disabilities by automatically magnifying or explaining words or context in the text (for example see [15,10]).

We propose Electrooculography (EOG) as a novel measurement technique for recognition of user activity and attention in wearable settings. EOG, in contrast to well established vision-based eye tracking, is measured with body-worn sensors, and can be implemented as a wearable system. Although requiring facial skin contact, we believe EOG electrodes

can be designed to be relatively unobtrusive, such as through integration into spectacles. A compact on-body device can then be used to process the incoming EOG signals.

The primary aim of this research is to assess the feasibility of recognising reading activity in different daily situations using wearable EOG. The wider goal of this is to gain insight into the potential of EOG for activity recognition. The specific contributions of the work are (1) an experiment involving data collection of subjects reading text while travelling to and from work, (2) a new method for saccade detection as a basis for reading recognition, and (3) an analysis of reading classification using string matching and Hidden Markov Models (HMMs).

The aim of our experiment is to capture reading in transit during different mobile situations. Despite the unavoidable fact that subjects wore sensing equipment on their faces, we took particular care to ensure that the chosen scenario - reading while travelling to and from work - was as realistic as possible. This scenario involved a continuous sequence of daily activities such as sitting at a desk, walking along a corridor, walking along a street, waiting at a tram stop and riding a tram. We recorded an 8 subject, ground truth annotated dataset, totalling nearly 6 hours of recordings - half of which involved reading.

Our work makes use of a new algorithm for detecting saccade features in EOG signals using Wavelet decomposition. Inspired by the typical characteristics of EOG signals during reading, we carry out a preliminary investigation into three different classification algorithms: a string matching algorithm on the horizontal saccade features; a discrete HMM also using the horizontal features; and a mixture of Gaussian HMM using the denoised signals from both horizontal and vertical EOGs. Our best result over all datasets was obtained using the string matching algorithm. Our main finding is that reading can be detected regardless of whether the subject is sitting, standing or walking, and in a variety of indoor and outdoor situations.

1.1 Related Work

In a recent work, Logan *et al.* aimed at recognising common activities in a "real world" setting using a large variety and number of common sensors such as wired reed switches, RFID tags and infra-red motion detectors in the environment [9]. They discovered that reading was one of the most difficult activities to detect and concluded that for covering all types of physical activity in daily life, additional sensors and improved algorithms need to be found.

All previous attempts to recognise reading have been based on vision to record eye movements. With the goal of building a more natural computer interface based on user activity, Campbell *et al.* investigated on-screen reading recognition using infra-red cameras to track eye movements [4]. The approach used was subject independent, robust to noise and had a reported accuracy of 100%. However, the system required that each subject's head was kept still using a chin rest.

In a later work, Keat *et al.* proposed an improved algorithm to determine whether a user is engaged in reading activity on a computer monitor [8]. Using an ordinary video camera placed between the subject and monitor, 10 subjects were asked to read an interesting text from a list of preselected articles. The subjects were explicitly asked to undertake other types of common computer-related activities such as playing computer games or watching video clips during the course of the experiment. Using user-dependent training, they achieved an average reading detection accuracy of 85.0% with a false alarm rate of

14.2%. However, to ensure correct detection of gaze direction, subjects were required to face the screen throughout the experiments.

Motivated by the goal of improving reading skills for people with reading disabilities, Sibert *et al.* developed a system for remedial reading instruction [15]. Based on visual scanning patterns, the system used visually controlled auditory prompting to help the user with recognition and pronunciation of words. Following the study, subjects reported that the most obtrusive part of the system was the video camera used to track eye movements.

Eye tracking using vision is a well studied field with a growing number of researchers looking at the movements of the eyes during daily activities in natural environments. Important advances are being made to the understanding of how our brains process tasks, and of the role that our visual system plays in this [7]. To assist with this work, a number of commercial trackers are available, some of which are targeted at mobile use. The most wearable of these - the Mobile Eye from Applied Science Laboratories (ASL) and the iView X HED from SensoMotoric Instruments (SMI) - both require bulky headgear and additional, cumbersome equipment to process the incoming video streams. To-date no solution for portable eye tracking exists that is convenient and unobtrusive enough to allow for unaffected physical activity.

Eye movement characteristics such as saccades, fixations and blinks, as well as deliberate movement patterns detected in EOG signals, have already been used for hands-free operation of static human-computer [13] and human-robot [23] command interfaces. All of these studies show that EOG is a promising measurement technique that can be remarkably accurate, easy to operate, reliable and can also be made cosmetically acceptable. Another interesting application is the use of EOG-based switches in a hospital alarm system which provide immobile patients with a safe and reliable way of signalling an alarm [19].

EOG-based interfaces have also been developed for assistive robots [22] and particularly as a control for an electric wheelchair [2]. These systems are intended to be used by physically disabled people who have extremely limited peripheral mobility but still retain eye motor coordination. Although both applications target mobile settings the peoples' movements are constrained and the situation therefore differs from the one investigated in this work.

For EOG to be truly unobtrusive - particularly for mobile settings - the design of novel electrodes and electrode configurations is a critical topic and still subject to research. Manabe *et al.* propose the idea of an EOG gaze detector using an electrode array mounted on ordinary headphones [11]. While this placement might reduce the problem of obtrusiveness, it raises two other issues - namely, low signal-noise ratio (SNR) and poor separation of horizontal and vertical components. In another work, Vehkaoja *et al.* made electrodes from conducting fibres and sewed them into a head cap [18]. As yet the device is still to be evaluated in operation.

2 Eye Movement Analysis

2.1 Wearable Electrooculography

The eyes are the origin of a steady electric potential field. This can be detected in total darkness and even while the eyes are closed. It can be described as a fixed dipole with its positive pole at the cornea and its negative pole at the retina. The magnitude of this corneoretinal potential (CRP) is in the range of 0.4-1.0 mV. It is not generated by excitable

tissue but is rather attributed to a higher metabolic rate in the retina. This potential difference is the basis for a signal measured between two pairs of surface electrodes placed in periorbital positions around one eye, the so-called Electrooculogram.

If the eyes move from the centre position towards one of these electrodes, the retina approaches this electrode while the cornea approaches the opposing one. This results in a change in the potential - the EOG signal - which can be used to track eye movements. The movement is split into horizontal and vertical signal components reflecting the discretisation given by the electrode setup.
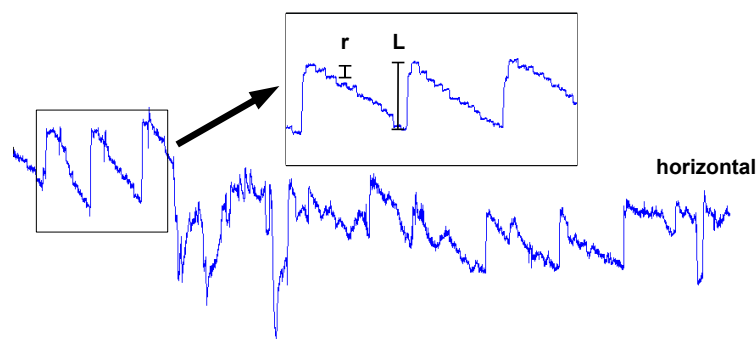

Fig. 1. Example signal from the horizontal EOG component with a reading segment in detail. The reading signal was denoised and baseline drift removed and shows typical large left (L) and small right (r) saccades occurring during reading.

2.2 Eye Movement Characteristics

To be able to take advantage of the typical characteristics of eye movements during reading, it is important to understand its two main types, namely saccades and fixations.

Saccades: Humans do not look at a scene in a steady way. Instead, their eyes move around and locate interesting parts of the scene to build up a mental "map" representing it. The main reason for this is that only a very small central part of the retina, the fovea, can sense in high resolution. The fovea has a very narrow field of view and in order to be able to see a wider scene, the eye must move constantly. The simultaneous movement of both eyes in the same direction is called a saccade. This is the fastest movement of any external part of the human body. The peak angular speed of the eyes during a saccade reaches up to 1000 degrees per second while lasting from about 20 to 200ms. The amplitude of a saccade is the angular distance that the eye needs to travel during a movement. For amplitudes of up to about 60 degrees, the duration of a saccade linearly depends on the amplitude. But beyond this, the velocity of the saccade remains constant and the duration of the larger saccades is no longer linearly dependent on the amplitude.

Eye movements during reading are characterised by a typical sequence of small and large saccades: First, sequences of small saccades occur while the eyes move over the words in a line of text. A large saccade is observed when the eyes move back to the beginning

of the next line of text. Figure 1 shows how both of these types of saccade look in the horizontal part of an EOG reading signal.

Fixations: Fixation is the static state of the eye during which gaze is held upon a specific location. Humans typically alternate saccadic eye movements and fixations. However, visual fixation is never perfectly steady and fixational eye movements can also occur involuntarily. The term "fixation" can also be referred to as the time between two saccades during which the eyes are relatively stationary. Reading involves fixating on successive locations within a line but also across a page to reach different sections of the text.

2.3   Baseline Drift

Baseline drift is a slow signal change mostly unrelated to the actual eye movements but superposing the EOG signal. Baseline drift has many possible sources as for instance interfering background signals, electrode polarisation [6] or physical influences such as varying contact pressure of the electrodes. In a four electrode setup as used in this study, baseline drift can also be different for the horizontal and vertical EOG signal components.

Baseline drift poses problems on the various types of eye movement: For saccades, the difference between start and end can be assumed to be drift-free, as saccades are performed in a very short period of time. Only signal segments before and after a saccade can become subject to changes caused by baseline drift. During periods of smooth pursuit movements it is difficult to distinguish baseline drift from actual eye movement; similarly for fixations drift alters the EOG signal in a way that can be indistinguishable from that of a slow eye movement.

Several approaches to remove baseline drift from electrocardiogram signals (ECG) have been proposed in recent literature (for example see [16,5]). As ECG shows repetitive characteristics, some of the algorithms perform sufficiently well at removing baseline drift from these signals. However, they perform worse for signals with non-repetitive characteristics such as EOG. Thus the development of robust algorithms for baseline drift removal is still an active topic of research.

3   Data Collection

The experimental setup devised in this work was designed with two main objectives in mind: (1) to record eye movements using EOG in an unobtrusive manner in a real-world setting, and (2) to evaluate how well reading can be recognised using EOG for persons in transit. We defined a scenario of travelling to and from work containing a semi-naturalistic set of reading activities. It involved subjects reading text while engaged in a sequence of activities such as sitting at a desk, walking along a corridor, walking along a street, waiting at a tram stop and riding a tram.

3.1   Experimental Procedure

Subjects were asked to follow two different sequences: A first calibration step involved walking around a circular corridor for approximately 2 minutes while reading continuously. The second sequence involved a walk and tram ride to and from work (see Table 3.1). This sequence was repeated in three runs: The first run was carried out as a baseline case

without any reading task. This was both to accustom the subjects to the route, but also to provide a reasonable amount of *NULL* data - which contributed to the objective of obtaining a realistic dataset. In two subsequent runs the subjects were asked to read a text throughout. Between each run the subjects rested for about 5 minutes. The total experiment time for each subject was about one hour. At the end of each experiment, the subjects were asked on their experiences on the procedure in a questionnaire.
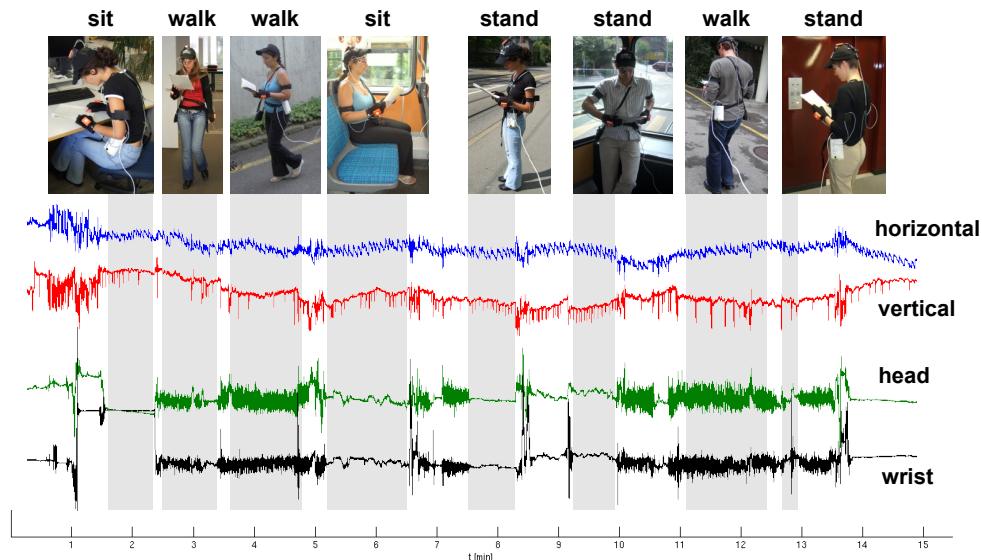

Fig. 2. Experimental procedure involving a sequence of semi-naturalistic reading activities including sitting at a desk, walking along a corridor, walking along a street, waiting at a tram stop and riding a tram. The figure also shows the corresponding horizontal and vertical EOG signals as well as acceleration data from the head and the right wrist of one complete dataset.

In contrast to a previous study [4] we opted to allow a free choice on reading material. Only two conditions were made: (1) that the material was text-only, i.e. no pictures and (2) that subjects only chose material they found interesting and long enough to provide up to an hour's worth of reading. Thus the type of text, its style as well as page and font size could be chosen to each subject's personal preference. Our objective was to induce a state where readers were engrossed in the task for the relatively long recording time, thus allowing us to gather realistic data without having to coerce subjects. A further benefit was that if subjects were engrossed in the task, they would be less likely to be distracted by other people.

We were able to collect data from 8 subjects - 4 male and 4 female - between the ages of 23 and 35. (Originally there were 10 subjects, but 2 had to be withdrawn due to recording problems resulting in incomplete data.) Most of the experiments were carried out in well lit, fair to cloudy conditions with two exceptions: One of the male subjects was recorded at night where we had to rely on street lights while walking around outdoors. Another male was recorded in rain where an assistant had to hold an umbrella over the subject

Average Time (min:sec)	Activity
00:00	Start synchronisation gesture
00:10	Sit at desk (on 3rd floor, indoors)
01:00	Walk to lift
01:10	Stand and wait for lift
01:40	Take lift to ground floor
02:10	Walk to exit
02:40	Walk to tram stop (outdoors)
04:40	Wait at tram stop
07:00	Ride the tram down a stop
08:00	Wait at tram stop (outdoors)
11:00	Ride the tram up a stop
12:00	Walk to entrance (outdoors)
14:00	Walk to lift (indoors)
14:30	Stand and wait for lift
15:00	Take lift up
15:30	Walk to office
16:00	Sit in comfortable chair
17:00	End synchronisation gesture

Table 1. Scenario of travelling to and from work. This was repeated 3 times for each subject: Once as a baseline measurement without reading and twice with subjects reading interesting texts. The total time per scenario averaged around 17 minutes: 6 min. indoors, 10 min. outdoors and 4 min. on tram).

to protect the sensors and reading material. However, as neither of the datasets showed a decrease in signal quality, both were used for the analysis.

Annotation of Ground Truth Each subject was tailed by an assistant who annotated both the subject's current activity (sitting, standing, walking) and whether he was reading. For this level of detail (is the subject's eyes on the page or not) the assistant had to monitor the subject from a close proximity - but without being so close as to cause a distraction. For this purpose we used a wireless controller from Nintendo, the Wii Remote (see Figure 3). Using the Wii's thumb control buttons "up", "down" and "right", the assistant could annotate the basic activities of standing, sitting and walking. In parallel, the trigger button was held down whenever the subject appeared to be reading and released when he stopped. A fifth button was used to annotate special events of interest, such as when the subject passed through a door and while entering or leaving the tram. All other buttons were not used and disabled to not interfere with the labelling.

Eye Movements For EOG data collection we used a commercial system, the Mobi from Twente Medical Systems International (TMSI), which was worn on a belt around each subject's waist (see Figure 3). The device is capable of recording a four-channel EOG with a joint sampling rate of 128Hz and transmitting aggregated data to a laptop carried by an assistant via Bluetooth.

The data was collected using an array of five electrodes positioned around the right eye (see Figure 3). The electrodes used were the 24mm Ag/AgCl wet ARBO type from
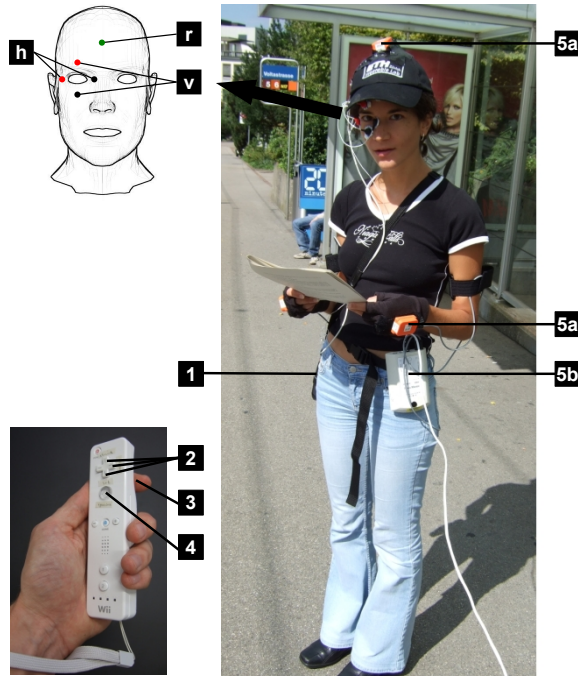
Fig. 3. Experimental setup consisting of five EOG electrodes (h: horizontal, v: vertical, r: reference) and the Mobi (1). The bottom figure shows the wireless Wii controller used for annotating the ground truth with three thumb control buttons (2), the reading trigger button (3) and a button for labelling special events (4). Also shown though not used in the current work are three Xsens motion sensors on the head and on the back of both hands (5a) and the XBus Master (5b).

Tyco Healthcare equipped with an adhesive brim to stick them to the skin. The horizontal signal was collected using one electrode on the nose and another directly across from this on the edge of the eye socket. The vertical signal was collected using one electrode above the eyebrow and another on the lower edge of the eye socket. The fifth electrode, the signal reference, was placed away from the other electrodes in the middle of the forehead.

Physical Activity One of our objectives for future work is to analyse the correlation of reading activities with physical movement and posture. Though beyond the focus of this current work, for completeness we include a short description of the additional recording setup. Briefly, this included MTx sensor nodes from XSens Technologies containing 3-axis accelerometers, magnetic field sensors and gyroscopes positioned on each subject's head as well as on the back of their hands (see Figure 3).

Unfortunately, the MTx system performed poorly using its bluetooth connection and so we were forced to use a wired connection. This was the only physical connection between the subject and assistant. Care was taken throughout by the assistant to ensure that the trailing wire did not interfere or distract the subject.

Data Recording All recorded data was sent to a laptop in the backpack worn by the assistant. Data synchronisation was handled using the Context Recognition Network (CRN) Toolbox (see [1] for details). We made two extensions to the toolbox: the first was a reader to process and synchronise labelled data from the Wii Remote controller; the second extension was to implement a "heartbeat" component that provided audio feedback to the assistant on whether the toolbox was running and recording data, thus providing instant notification of device failures. The addition of the "heartbeat" was particularly useful as it allowed the assistant to concentrate on labelling and observing the subjects rather than continually disturbing the procedure by checking the recording status.

4   Continuous Recognition Methods

The methods in this work were all implemented offline using MATLAB. However, with a view to a future online implementation on a wearable device, the algorithms were chosen to keep computation costs low. In this section we first describe the signal processing steps required for removing noise and baseline drift, and for extracting saccade information. We then describe three algorithms we use for classification: string matching (STR), discrete HMM (D-HMM) and mixed Gaussian HMM.

Noise and Baseline Drift Removal In a parallel work we evaluated algorithms for baseline drift removal and integrated them into a general artefact compensation framework for EOG signals [3]. There, we adapted a wavelet packet approach originally proposed for ECG signals [16]. First, the data is stripped of high frequency noise using a simple median filter. The algorithm then performs an approximated multilevel 1-D wavelet decomposition at level nine using Daubechies wavelets on the horizontal signal component. The reconstructed decomposition coefficients then give a baseline drift estimation. Subtracting this estimation from the original signal finally yields the corrected signal with reduced drift offset.

Saccade Detection and String Encoding Both the string matching and the D-HMM algorithms rely on the detection of saccades in the horizontal component of the reading signal. For this purpose we developed the so-called *Continuous Wavelet Transform - Saccade Detection* (CWT-SD) algorithm, and applied it to the de-noised, baseline drift removed, horizontal EOG signal [3]. The CWT-SD first computes the continuous 1-D wavelet coefficients from the signal at scale 20 using Haar wavelets. A saccade is detected for all samples where the absolute value of the corresponding coefficient vector exceeds a threshold.

To keep the algorithm simple, the string encoding for this initial work only uses the horizontal component of the signal[3]. The saccade detection is run twice using two different threshold values: T_{sac}^s represents small saccades, such as jumps between words during reading; and T_{sac}^L represents large saccades, such as those observed during an end of line "carriage return". The resulting sequence of large and small saccades together with their direction are then encoded into a string using characters according to the following scheme:

- "**L**": large saccade to the left
- "**R**": large saccade to the right

---

[3] Our future work will be based on 2D encoding using both horizontal and vertical information.

– **"l"**: small saccade to the left
– **"r"**: small saccade to the right

An example of an encoded reading segment is shown in Figure 4.
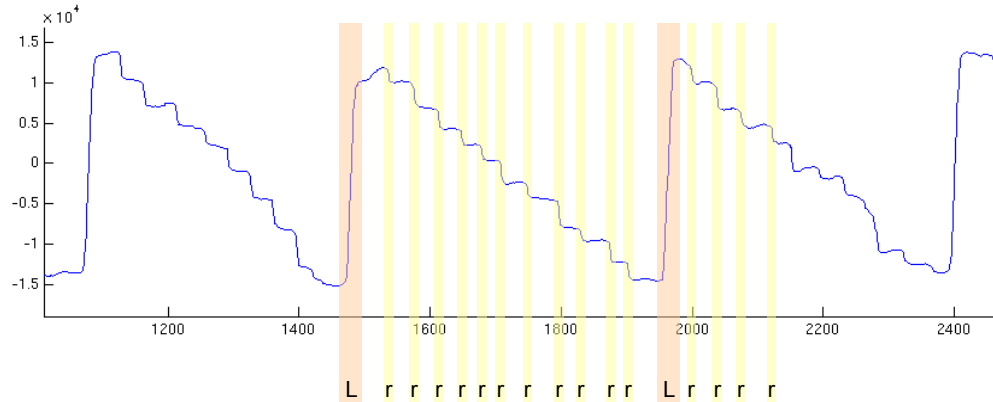

Fig. 4. Horizontal EOG reading signal and corresponding string encoding

4.1 String Matching

Of all the algorithms presented here, string matching is computationally the most simple. It can be considered a light-weight approach, using only simple arithmetic, and can be easily adapted to a future online implementation, for example on a wearable device.

The matching is performed by moving a prototype string, representing a typical reading segment, over the signal string encoding, character by character. In each of these steps, the Levenshtein distance between the string template and the current signal string is calculated. The Levenshtein distance between two strings is given by the minimum number of operations needed to transform one string into the other, where an operation is an insertion, deletion, or substitution of a single character.

The algorithm then applies a threshold T_{ed} on the Levenshtein distance vector to separate the two classes "reading" and "not reading". This threshold defines how tolerant the classification is towards relative error in the edit distances. As this method does not yet adapt the string template to the signal while calculating the distances, it is sensitive to fluctuations in the number of small saccades. This results in a high number of false insertions. To counter this we slide a majority vote window W_{str} across the event-based output classification to "smooth" the final result.

4.2 Hidden Markov Models

Hidden Markov Models are probabilistic models used to represent non-deterministic processes in partially observable domains and are defined over a set of states, transitions and

observations. In this work we evaluate two different implementations of HMM: discrete and mixture of Gaussian. In each we use a 2-state model. This helps to keep the computation costs low. Intuitively, the two states represent the large left saccades when the eyes move back to the beginning of a line and the saccades during word reading. The parameters - state transition probabilities, observation likelihoods and Gaussian mixture settings - are set in a training step using the Baum-Welch (*forwards-backwards*) algorithm. For classification, we recursively apply the *forward* algorithm on incoming data samples. See [14] for details on these algorithms, and [12] for the Matlab implementation we used.

HMMs with Discrete Observations For the discrete case, we use the same character features as for string matching. First we segment the incoming data using the large left saccades as boundaries (with each saccade beginning with an "L"). This leaves us with a sub-string of saccades ("r","l" and "R") for each segment, which we then feed into the discrete HMM. Note that in this simplified model, the influence of a large right saccade, "R", which might be encountered during page turns, is not modelled explicitly. For each segment, the *forward* algorithm returns a single log likelihood value. A threshold T_{dhmm} is then applied and if the threshold is passed, the classifier returns reading, if not *NULL* is returned.

HMMs with Gaussian Observations For the Gaussian case, we use the horizontal and the vertical EOG signal component as the observation feature space (de-noised and baseline drift removed). To avoid singularities with the implementation of the *forward-backward* algorithm, we standardise the signals by setting the variance of the entire dataset to 1. Our model is again 2 state, but with a mixture of two Gaussian probability density functions to model the observations.

At each of its steps, the *forward* algorithm outputs a log likelihood value. We then smooth the sequence of these likelihoods using a sliding mean window W_{hmm}. Finally, a threshold T_{hmm} is applied and if the threshold is passed, the classifier returns reading, if not *NULL* is returned.

5   Results

5.1   Parameter Selection and Training

Saccade Detection and String Matching To determine the threshold parameters we applied the saccade detection algorithm using a threshold sweep on a manually cut subset of the data. On average for all subjects, this subset of data contained 15 large reading saccades plus noise and artifacts caused by interrupting eye movements. For each threshold, we counted the number of large saccades that were detected and calculated the *relative error* ($\frac{Total-Detected}{Total}$). Based on this sweep we chose the large saccade threshold at $T_{sac}^{L} = 7000$. Due to the difficulties in manually segmenting samples of small saccades, we approximated the small threshold $T_{sac}^{s} = 1500$.

The string matching parameters were evaluated across a sweep of the majority vote window length W_{str}, the distance threshold T_{ed} and a selection of different templates. In the analysis presented below, we chose to fix $W_{str} = 30$ and template "*Lrrrrrrrr*" for all subjects. Figure 5 shows an example output of the basic algorithm compared against the smoothed result and the ground truth labelling. Note the typical errors encountered:
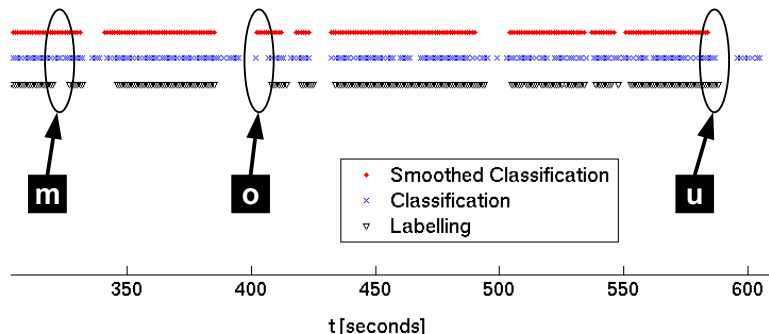
Fig. 5. Example string matching result showing the ground truth labelling, the classification result returned by the algorithm and the result after applying the smoothing filter. Also shown is an example merge (m) and an overfill (o) and underfill (u) error.

"merge" (m) when the output detects a single reading sequence where there should be two; "overfill" (o) when the output correctly detects reading just outside the labelling; and "underfill" (u) where the detected reading signal falls short of the labelling. These error types are explored in more detail later in this section.

HMM Training Both HMM and D-HMM models were trained using data from the recordings of reading while walking around a corridor. Two leave-one-out training schemes were used: subject dependent only using the calibration data from the subject being tested and subject independent using calibration data only from other subjects.

We evaluated the HMMs over a sweep of the main parameters. For the mixed Gaussian HMM, we discovered that sliding window size had limited influence, and so could be fixed at $W_{hmm} = 5$ seconds.

5.2 Continuous Classification

The different methods were compared across a sweep of their main parameters: $T_{ed} = 1...10$ (in 10 steps), $T_{hmm} = T_{dhmm} = -1...-9$ (in 32 steps). The results from all subjects were then summed together. The resulting Receiver Operating Characteristics (ROC) curves are shown on the left of Figure 6. These plot true positive rate (recall) ($\frac{TP}{TP+FN}$) against false positive rate (FPR) ($\frac{FP}{FP+TN}$), where TP, FP, TN and FN represent true positive, false positive, true negative and false positive counts respectively. Best case results approach the top left corner while worst case (which means random) follow the diagonal.

The ROC clearly shows that string matching outperforms the HMMs. At its "best" ($T_{ed} = 5$), string matching returns a recall of 71.0% and FPR of 11.6% (total accuracy 80.2%). The mixed Gaussian returns a lower best-case at recall 62.0%, FPR 24.0% and accuracy 68.9% for $T_{hmm} = -3.5$ and subject independent training. The worst performing algorithm is the discrete HMM. However, both types of HMMs perform similar for the subject-dependent and the subject-independent case.
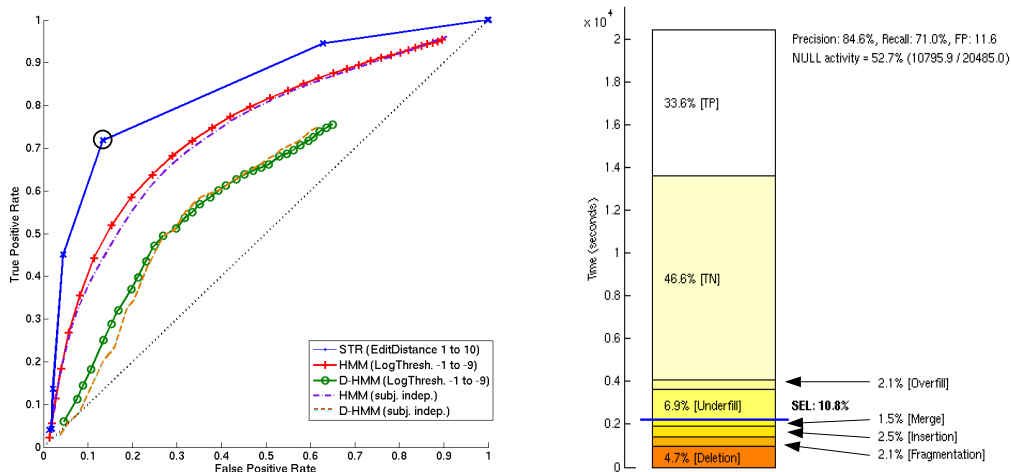
Fig. 6. Left: ROC curves showing a performance comparison between string matching (STR) over a sweep of edit distance T_{ed}, Gaussian HMMs over a sweep of log threshold T_{hmm} and discrete HMM over a sweep of T_{dhmm}. For the HMMs, both the subject-dependent and the subject-independent results are shown. Right: Detailed result for string matching with $T_{ed} = 5$ (corresponding to the point circled on the ROC curve). The EDD shows the proportion of the total dataset comprising true positives (TP), true negatives (TN), overfill, underfill, merge, insertion and fragmentation errors. Note that the proportion of negatives ($NULL$) in the dataset is roughly half (52.7%). Though overall accuracy is 80.2% ($TP+TN = 33.6\% + 46.6\%$), a large part of the errors are actually underfill and overfill timing errors.

Further Analysis of String Matching Based on these results, we chose string matching (with $T_{ed} = 5$) for further analysis. The error division diagram (EDD) of Figure 6 shows a detailed timewise breakdown of the errors. The EDD highlights typical errors that occur in continuous recognition systems - beyond the basic FP and FN categorisation. Specifically there are three classes of error that we consider. Details on how these error classes are derived is outwith the scope of the current work, but can be found in [20]:

1. The "classical" errors, such as insertion (a reading event is detected where there is none in the ground truth) and deletion (failure to detect a reading event).
2. Fragmentation and merge: *Fragmentation* errors describe when a reading event in the ground truth corresponds to several events in the recognition system output. *Merge* is the opposite: several ground truth reading events are combined into one event - see (m) in Figure 5.
3. Timing errors: *Overfill* errors are where an event in the output of the system extends into regions of $NULL$, such as in the event (o) in Figure 5. The opposite of *overfill* is *underfill* (u), in this case the event recognised by the system fails to "cover" some parts of the ground truth event.

In addition, EDDs also show the total true negative (TN) and true positive (TP) times. Using this breakdown, Figure 6 shows that 9.0% of the total time is underfill and overfill "error". These are cases where the fault may be slightly offset labelling, or delays in the recognition system. The errors that might be regarded as more serious - insertion,

deletion, merge and fragmenting - account for 10.8% of the total experiment time, this is the so-called serious error level (SEL).
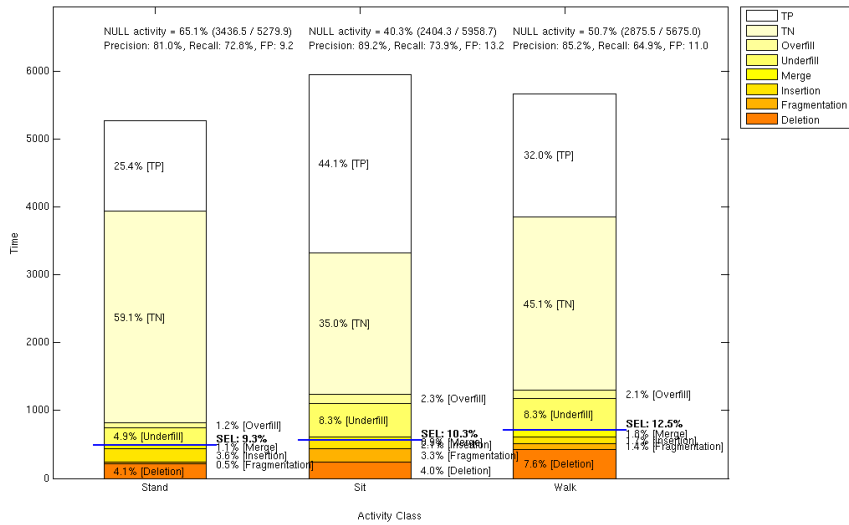

Fig. 7. Performance evaluation for reading recognition during distinct activities - stand, sit and walk - using the string matching algorithm. The EDD shows the proportion of the total dataset comprising true positives (TP), true negatives (TN), overfill, underfill, merge, insertion and fragmentation errors. Note the different distributions of *NULL* in each of these cases: 65.1%, 40.3% and 50.7% for stand, sit and walk respectively. As expected, the results for walking are slightly worse than the others, with 12.5% serious error level (SEL). Interestingly, the sit case is slightly worse than stand. This is due in part to the greater number of fragmentation errors from sit.

Reading in Different Situations To analyse the performance of string matching for the different activity situations, we divided the data into three sets, separating sitting, standing or walking activities. Each activity represented roughly equal sized portions of the dataset (88, 98 and 95 minutes respectively). The EDDs in Figure 7 show the results of this evaluation over all subjects.

For standing, the best case result shows 72.8% recall and 9.2% FPR; for sitting the result is 73.9% recall, 13.2% FPR; and for walking it is 64.9% recall and 11.0% FPR. Looking further at the type of errors encountered, we see that both sit and walk contain large periods of underfill timing errors. Also of interest is the fact that 3.3% of the time during sit is classed as a fragmentation error.

Results for Each Subject The results for each individual subject show a range of differences in recognition performance using the string matching algorithm (see Table 5.2). The highest recall result is 92.9% (subject 3) but with a FPR of 15.5%. The worst result was for subject 1, with 25.9% recall and 3.8% FPR. This was to be expected as

	S1 (f)	S2 (m*)	S3 (f)	S4 (m**)	S5 (f)	S6 (m)	S7 (f)	S8 (m)	median
Recall [%]	25.9	65.2	92.9	83.1	47.9	89.5	71.3	85.7	**77.2**
FPR [%]	3.8	12.1	15.6	17.9	4.3	18.3	12.9	8.9	**12.5**

Table 2. Recall and FPR for each individual subject, and the median over all, using string matching. The table also shows the subjects' gender (f: female, m: male); the dataset recorded at night is marked with * while the one recorded in heavy rain with **.

the raw EOG signal quality for that subject was extremely low. During the experiment recordings, the overall signal level for this subject was known to be weak and reading saccades could hardly be seen. This was also a problem for subject 5 and was most likely caused by poor electrode placement and dry skin.

What can be seen from the table, however, is that the differences do not seem to correlate to the gender of the person. Also the two special datasets recorded at night and in rain do not show marked differences, but have comparable performance to the other sets.

6 Discussion

On the recognition performance Three different approaches of recognising reading based on EOG in a wearable setting have been described and investigated in this work. Among the algorithms evaluated, string matching performs best, with an overall recognition rate of 80.2% (71.0% recall, 11.6% FPR). It is interesting to note that almost half of all errors are overfill and underfill. The most probable causes of these errors are the inaccuracies of the sliding majority filter and the labelling process.

The D-HMM method does not perform well. This is probably due the lack of sufficiently descriptive features. We see in Figure 6 that it is robust to different training configurations. This indicates that the novel method of saccade detection CWT-SD, upon which both string matching and D-HMM methods are based, is fairly robust to variations in EOG signals. This is a particularly useful trait for applications where a "training" step would be inconvenient or impractical. A further advantage of the CWT-SD is that it only requires the horizontal EOG signal component and therefore only 2 electrodes. Although we strongly believe that most of the useful information lies in the horizontal component, the (as yet unused) vertical component might also contain information that in a future study could improve the performance of both methods.

Despite using both horizontal and vertical signal components and the perceived aptitude of HMMs for this type of problem, the Gaussian HMM did not produce the results that have been expected. Part of the reason for this relatively poor performance (compared to the simple string matching) might be due to the specifics of our initial implementation. It is the authors' belief that with careful selection of more descriptive features and perhaps a larger number of states, this method can be improved.

By analysing the different activity situations - reading while standing, sitting and walking - we uncover some interesting results. As expected, walking produces the worst results but only by a small margin - its serious error level is only 2.2% worse than sitting (Figure 7). Surprisingly, recognition is better while subjects stand than when they sit. We would have expected the "sit" class to perform best as this is the one with potentially the

least external influences. So far, we have no satisfactory explanation for this behaviour but we plan to investigate this further in detail in future experiments.

On EOG From this initial study we found that EOG is a robust technique for recording eye movements in wearable settings. The main advantage of the EOG-based measurement technique is the fact that, in contrast to common video-based eye trackers, the subject only has to wear relatively unobtrusive and light-weight equipment. This contributes to the subjects feeling unconstrained during the experiments and therefore allows for natural reading behaviour and unaffected physical activity.

One drawback is that EOG electrodes require good skin contact. Poor placement of electrodes was the reason for many of the problems in our work, particularly with subjects 1 and 5 (see Table 5.2). This problem was usually solved by removal and reattachment of fresh electrodes. The fact that these electrodes are stuck to the face can also be regarded as inconvenient. In the post-experiment questionnaire, the subjects from our study reported that they did not feel physically constrained by the sensors and wires - not even by the electrodes. However, it is clear that for long-term use a more comfortable and robust solution would be desirable.

Baseline drift is perhaps an unavoidable problem for wearable EOG recording. It is for this reason that accurate gaze tracking, for purposes such as target detection, might be difficult to achieve using mobile EOG. By analysing only the rough patterns created by eye movement, however, we can detect activities (such as reading) without the need for such pinpoint tracking.

On the experiment The Wii-remote proved to be a useful and unobtrusive annotation tool - and was certainly preferable to the chore of video-based offline annotation. This method is certainly subject to inaccuracies when, for example, the assistant is distracted, or when buttons are pressed and released too early or too late. However, labelling errors are an intrinsic problem especially in wearable settings and a satisfying solution has not been found yet. One possible approach for a future study could be to introduce redundancy to the labelling process: either a second assistant, or some semi-automatic solution, perhaps using additional sensors (e.g. using body movements or video-based eye tracking).

In the questionnaire, all subjects declared that they did not feel distracted by people on the street and were only partially conscious about the experiment assistant. Half of the subjects did report a feeling of unease while reading and walking. This unease could clearly be seen in the EOG signal by the occasional presence of small vertical saccades during reading - indicating whenever a subject looked up to check the way ahead.

Obviously reading while walking can be a dangerous activity, however this does not detract from the fact that many people actually do it. The other half of our subjects found no problem with reading and walking - they all claimed to have occasionally read long texts (a book, newspaper or scientific paper) while in transit. It is more common for people to read shorter texts while walking, for example advertisements, timetables, etc., and this would be an interesting scenario for future study. Because such reading sequences are usually very short, the recognition methods and also the labelling scheme would probably have to be adapted.

Ideally, the most natural scenario would have involved recordings over a period of weeks or months. This would allow us to better study the general reading behaviour of our subjects - and to open up interesting questions regarding daily reading habits.

Unfortunately, the battery life and reliability of our recording equipment limited recordings to a few hours. Therefore, the main improvement concerning the experimental setup is to develop a wearable EOG device which does not impose these restrictions but allows for robust long-term eye movement recordings. This would also include an investigation of how to apply dry electrodes, e.g. by integrating them into spectacle frames, as they are more convenient for everyday use than wet electrodes stuck to the skin. Another interesting question is how dry electrodes would influence the quality of the signals and recognition performance.

6.1 Conclusion

Our work has shown that wearable EOG is a feasible approach for recognising reading in daily-life scenarios and is robust across an example set of activities for different subjects. This raises the question of whether different reading behaviours and attention levels to written text can be detected automatically. A "reading detector" could enable novel attentive user interfaces which take into account aspects such as user interruptability and level of task engagement.

Given appropriate hardware, EOG offers the potential of long-term eye movement recordings. The movement patterns the eyes follow in daily routine reveal much about what people are doing - as well as what they intend to do. With growing interest in activity recognition as a topic within pervasive computing, this information may prove extremely relevant. In the future, eye movements may be used as a new sensing modality, providing access to the underlying cognitive processes not available with current sensing modalities.

7 Acknowledgements

References

1. D. Bannach, K. Kunze, P. Lukowicz, and O. Amft. chapter Distributed Modular Toolbox for Multi-modal Context Recognition, pages 99–113. 2006. 10.1007/11682127_8.
2. R. Barea, L. Boquete, M. Mazo, E. Lopez, and L. M. Bergasa. EOG guidance of a wheelchair using neural networks. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 4, pages 668–671, Barcelona, Spain, 2000.
3. A. Bulling, P. Herter, M. Wirz, and G. Tröster. Automatic artefact compensation in EOG signals. In *EuroSSC 2007: Adjunct Proceedings of the 2nd European Conference on Smart Sensing and Context, October 23-25*, 2007.
4. C. S. Campbell and P. P. Maglio. A robust algorithm for reading detection. In *PUI '01: Proceedings of the 2001 workshop on Perceptive user interfaces*, pages 1–7, New York, NY, USA, 2001. ACM Press.
5. V. S. Chouhan and S. S. Mehta. Total removal of baseline drift from ECG signal. In *Computing: Theory and Applications, 2007. ICCTA '07. International Conference on*, pages 512–515, March 2007.
6. J. J. Gu, Max Meng, A. Cook, and M. G. Faulkner. A study of natural eye movement detection and ocular implant movement control using processed EOG signals. In *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on*, volume 2, pages 1555–1560, 2001.

7. M. Hayhoe and D. Ballard. Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9:188–194, Apr 2005.

8. F. T. Keat, S. Ranganath, and Y. V. Venkatesh. Eye gaze based reading detection. In *TENCON 2003. Conference on Convergent Technologies for Asia-Pacific Region*, volume 2, pages 825–828, October 15–17, 2003.

9. B. Logan, J. Healey, M. Philipose, E. Tapia, and S. Intille. chapter A Long-Term Evaluation of Sensing Modalities for Activity Recognition, pages 483–500. 2007. 10.1007/978-3-540-74853-3_28.

10. P. P. Maglio, T. Matlock, C. S. Campbell, S. Zhai, and B. A. Smith. chapter Gaze and Speech in Attentive User Interfaces, page 1. 2000.

11. H. Manabe and M. Fukumoto. Full-time wearable headphone-type gaze detector. In *CHI '06: CHI '06 extended abstracts on Human factors in computing systems*, pages 1073–1078, New York, NY, USA, 2006. ACM Press.

12. K. P. Murphy. The HMM toolbox for MATLAB, http://www.ai.mit.edu/ murphyk/software/hmm/hmm.html, 1998.

13. D. Qiuping, T. Kaiyu, and L. Guang. Development of an EOG (electro-oculography) based human-computer interface. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pages 6829–6831, September 01–04, 2005.

14. L. R. Rabiner and B. H. Juang. An introduction to hidden Markov models. *IEEE ASSP Magazine*, pages 4–16, Jan 1986.

15. J. L. Sibert, M. Gokturk, and R. A. Lavine. The reading assistant: eye gaze triggered auditory prompting for reading remediation. In *UIST '00: Proceedings of the 13th annual ACM symposium on User interface software and technology*, pages 101–107, New York, NY, USA, 2000. ACM Press.

16. M. A. Tinati and B. Mozaffary. A wavelet packets approach to electrocardiograph baseline drift cancellation. *International Journal of Biomedical Imaging*, 2006:Article ID 97157, 9 pages, 2006. doi:10.1155/IJBI/2006/97157.

17. K. Van Laerhoven, H.-W. Gellersen, and Y. G. Malliaris. Long-term activity monitoring with a wearable sensor node. *bsn*, 0:171–174, 2006.

18. 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 *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pages 5865–5868, September 01–04, 2005.

19. S. Venkataramanan, P. Prabhat, S. R. Choudhury, H. B. Nemade, and J. S. Sahambi. Biomedical instrumentation based on electrooculogram (EOG) signal processing and application to a hospital alarm system. In *Intelligent Sensing and Information Processing, 2005. Proceedings of 2005 International Conference on*, pages 535–540, January 4–7, 2005.

20. J. A. Ward, P. Lukowicz, and G. Tröster. Evaluating performance in continuous context recognition using event-driven error characterisation. In *Location- and Context-Awareness*, pages 239–255, 2006. 10.1007/11752967_16.

21. J. A. Ward, P. Lukowicz, G. Tröster, and T. E. Starner. Activity recognition of assembly tasks using body-worn microphones and accelerometers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10):1553–1567, October 2006.

22. 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 *Robotics and Biomimetics (ROBIO). 2005 IEEE International Conference on*, pages 490–494, 2005.

23. C. Yingxi and W. S. Newman. A human-robot interface based on electrooculography. *2004. Proceedings. ICRA '04. 2004 IEEE International Conference on Robotics and Automation*, 1:243–248, April 26–May 1, 2004.