# Quantifying Reading Habits – Counting How Many Words You Read

**Kai Kunze, Katsutoshi Masai, Masahiko Inami**
Keio Media Design
Yokohama, Japan
Email:{kai/masai/inami}@kmd.keio.ac.jp

**Ömer Sacakli, Marcus Liwicki, Andreas Dengel**
DFKI
Kaiserslautern, Germany
Email:faruksacakli@gmail.com, liwicki@dfki.uni-kl.de, andreas.dengel@dfki.de

**Shoya Ishimaru, Koichi Kise**
Osaka Prefecture University
Sakai, Japan
Email:shimaru@m.cs.osakafu-u.ac.jp, kise@cs.osakafu-u.ac.jp

## ABSTRACT
Reading is a very common learning activity, a lot of people perform it everyday even while standing in the subway or waiting in the doctors office. However, we know little about our everyday reading habits, quantifying them enables us to get more insights about better language skills, more effective learning and ultimately critical thinking. This paper presents a first contribution towards establishing a reading log, tracking how much reading you are doing at what time. We present an approach capable of estimating the words read by a user, evaluate it in an user independent approach over 3 experiments with 24 users over 5 different devices (e-ink reader, smartphone, tablet, paper, computer screen).We achieve an error rate as low as 5% (using a medical electrooculography system) or 15% (based on eye movements captured by optical eye tracking) over a total of 30 hours of recording. Our method works for both an optical eye tracking and an Electrooculography system. We provide first indications that the method works also on soon commercially available smart glasses.

### Author Keywords
Mobile Eye tracking; Electrooculography; Quantifying Reading; Eye Movement Analysis; Reading Behavior

### ACM Classification Keywords
H.5.2 Information interfaces and presentation (e.g., HCI): Miscellaneous.

## INTRODUCTION
Increased reading volume is associated with numerous cognitive benefits, including improved vocabulary skills, higher general knowledge and increased critical thinking [17]. Furthermore, reading is entertaining and has social value, higher reading volumes in adolescents are correlated with higher self-esteem and improved cognitive and emotional well-being[38]. Although there are these strong positive effects, only few previous works evaluated reading activities in situ and even fewer tried to quantify them[9, 35].

Despite the growing awareness on how important reading is for learning, it's challenging to get people to read more, especially as the amount easy digestible content in form of videos etc. increases. Automatically tracking physical activities can motivate users to more healthy lifestyles [6]. We believe this translates also to cognitive skills and tasks, as students can already boost their learning rate by keeping a manual record of their activities [27]. We want to investigate wether we can track reading habits similar to physical activity to give users tools to improve their mental fitness. Letters and words represent ideas and concepts; tracking the volume, speed and time a user is reading them seems particularly valuable as it gives first insights into learning and provides us with a basic countable measure of our performance [25]. For example, children suffering from reading disabilities can be earlier diagnosed, people can without trouble improve their reading speed and older adults have an easier way to fight dementia. Since research suggests that performance related to these situations is closely linked to reading volume[42, 26, 41].

We still have a hard time defining what healthy reading habits for adults are[17], as tools are missing to quantify reading in everyday situations and in long term studies. This paper provides the first steps towards assessing reading volume in realistic settings utilizing mobile eye tracking.

The contributions of this work are: (1) We present methods to quantify how much words a user reads working for both, optical eye tracking and electrooculography, (2) we achieve the lowest error rate between 5 -15% user-independent for our word count estimation over 3 data sets of in total 24 users with over 30 hours of eye gaze recordings, (3) we provide initial evidence that our methods to estimate word count can work on consumer smart glasses, performing a small user study on a first EOG glasses prototype.

## RELATED WORK

### Alternative Modalities
As we are interested in tracking reading habits, a collection of cognitive tasks, we first might try direct brain sensing. Yet, as related work shows most methods are too bulky or are quite noisy to get decent results related to recognizing reading [36,


*UbiComp '15*, September 7-11, 2015, Osaka, Japan.

ACM 978-1-4503-3574-4/15/09 . . . $15.00
http://dx.doi.org/10.1145/2750858.2804278

16, 20, 18]. The pre-processing steps seem more computationally complex compared to EOG or optical eye tracking.

The most interesting modalities for direct brain sensing seem to be electroencephalography (EEG) and near-infrared spectroscopy (NIRS), as both can be used in mobile settings[48, 46, 22]. However, for both spacial and for NIRS temporal resolution are not so good. Their signal is also strongly affected by motion noise and usually requires more complex filtering/pre-processing steps compared to EOG/optical eye tracing [9, 35].

### Eye tracking

However, the strong relationship between reading and eye movements is very well explored in cognitive science and psychology [45, 29]. For example, Kligel et al. investigate correlations of eye fixations with cognitive tasks related to reading [32]. Rayner provides a good summary of eye tracking research [43]. Most of the reading research in psychology however emphasizes on older adults or disabled [20, 16].There are only a few research publications centering around reading detection in mobile and stationary settings [10, 9]. Such reading detection algorithms can be used as a very simple word counting mechanism, as there's a relation between time read and the read volume. Biedert et al. look into how people read text. They presented a method to discriminate skimming from reading using a novel set of eye movement features [5]. Their algorithm works in real-time, deals with distorted eye tracking data and provides robust classification accuracies of 86% accuracy. They also showed a method to recognize text comprehensibility with an accuracy of 62% from gaze data recorded from multiple readers [4]. Buscher et al. proposed eye movements as automatic relevance feedback for information retrieval tasks[13].

### Enhancing the Reading Experience

In a series of works, Biedert et al. studied ways to enhance the reading experience of the user. They presented EyeBook [2] and Text 2.0 [3] a reading interface that observe which part of the text is currently being read by the user and that generate appropriate effects (e.g. playing sounds). However, they don't evaluate what suitable interventions are to increase users enjoyment, comprehension or attention. Xu et al. apply eye movement analyzes for document summaries, yet the environment is very controlled, e.g. the users need to rest their chin on a support when performing the reading task [49].

Concerning reading habits, there are some questionnaire based evaluations giving advice about effective reading techniques to second language learners, as well as for children with reading disabilities and older adults struggling with dementia[21, 42, 41]. Hansen [26] reports on a series of studies on reading comprehension with rapid readers trained in the Evelyn Wood method. There are also a couple of other works exploring speed reading and comprehension[45, 19, 15, 24], giving advice about reading techniques to increase speed and understanding. Several mention rigorous practice and steady increase in reading volume as one of the key factors to success [31]. There are also a couple of papers exploring speed reading together with eye movement analysis[31, 29]. Busher et al. discuss in general the feasibility of gaze based annotations for documents [14].

### Cognitive Task Tracking

There are also some efforts to infer the users expertise, language skill and other higher level cognitive activities using eye tracking [34, 37, 11, 23, 7, 28]. Most of the research focusing on second language learners or infants as improvements can be easier tracked using indirect measurements (questionnaires etc.). Bulling et al. coin the term cognition-aware computing to describe computing able to understand and support our mental activities not focusing alone on eye tracking but brain sensing in general [11]. Orchard et al. try to assess other cognitive states, especially cognitive workload, while users read by analyzing blinking patterns[40]. Rudmann et al. give a concise overview about cognitive state detection using visual behavior [44].

### Toward Quantified Reading

The closest to our work is the Wordometer implemented by Kunze et al. [35, 33]. They introduce word counting algorithms also based on mobile eye tracking and EOG (however both algorithms are separate). For the optical systems, their work relies on document image retrieval for filtering and mapping the eye gaze into the coordinate system of the read text. Therefore, they need to use the scene camera of the eye tracker (not only for document identification, but also for eye gaze filtering). We see our work complementary and as an extension of their method. As our comparison with their algorithm shows we improve the average error rate from 45 % to 8% for a complex dataset. Their mobile eye tracking method cannot cope with varying line lengths often found on different reading media (e.g. tablet versus news paper). In addition, all of their participants were Japanese and the experimental setup was more constrained. We in turn present a dynamic line break detection as well as a novel reading detection and word count estimation based on saccade features only, making the whole method more versatile and portable to other eye tracking or eye movement analysis systems.

As far as we know, this is the only research work exploring technology support to quantify reading and presenting a word count estimation algorithm capable of dealing with varying device types, line lengths working for both common eye tracking techniques.

## APPROACH

As seen from the related work section, using eye motions seem to be a promising approach to quantify reading habits. There are two common techniques for tracking eye gaze/motion: electrooculography (EOG) and optical tracking. EOG uses electrodes to measure a change in potential when the eye moves, as the eye can be represented as a dipole between the cornea and retina, This approach is cheap to implement and requires little processing compared to optical tracking. However, it just gives relative eye movements. Alternatively, we can track eye gaze using cameras and infrared light, providing potentially higher accuracy but requiring more processing power, inferring eye position, motion and gaze based on iris shape.

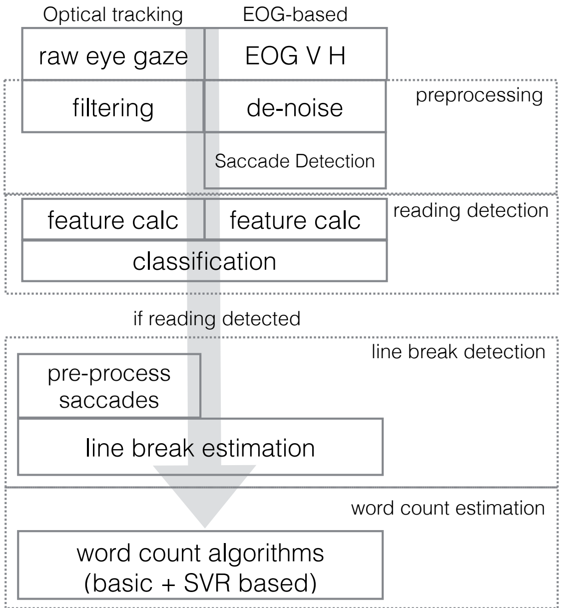



Figure 1: Approach overview, on the left for optical and right EOG systems.

Our method works in principle with both most of these techniques with slight adaptations. Our approach is divided into 4 discrete steps: preprocessing, reading detection, line-break detection, word count estimation. However, some of the steps are specific to the given technique also highlighted in the method overview as seen in Figure 1(left optical eye tracking, right EOG based eye movement analysis).

We get either the raw eye gaze data from the (mobile) optical eye tracker (fixation and saccade information), or horizontal/ vertical component from the Electrooculography. We use an EOG setup that can be integrated in smart glasses shown later Figure 7, similar to Kunze et al. [33].

### Preprocessing

*Optical Eye Tracking* –We combine several small, close-by fixations into larger duration fixations using the method of Busher et al. [12].

*EOG* – We apply a Median filter using a sliding window of size $w$ to filter noise. For saccade detection we use the Continuous Wavelet TransformSaccade Detection (CWT-SD) described Bulling et. al.[10]. We don't perform the letter encoding of CWT-SD, just use it to get saccade direction and amplitude.

### Reading Detection

The method is straight forward. We calculate the common features given in Table 1 over a 3 sec. frame sliding window and apply a Support Vector Machine classifier with a radial basis function on the resulting feature vector.

For the optical system we additionally calculate mean fixation duration and variance of the fixation count over the sliding window. For EOG we add blinking duration and frequency

| | |
|---|---|
| saccade related features | average length of saccades<br>minimum length of saccades<br>horizontal element of saccades<br>vertical element of saccades<br>saccade direction mean and variance<br>saccade slope |

Table 1: Common features for reading detection

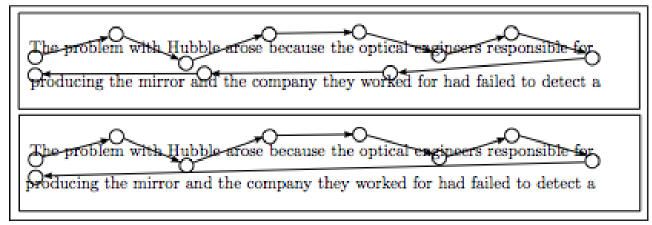



Figure 2: Preprocessing for Line break detection. Top: unprocessed eye gaze while reading. Bottom: Processed Eye gaze, saccades against the reading direction are combined.

(the blink detection algorithm is similar to Bulling et. al. [10]) as features.

As already mentioned in related work, methods for reading detection are not new [5, 8]. However, some of them work using different sensing modalities and eye tracking hardware. We wanted to present a whole system. Additionally, the reading detection method can be seen as a very simple word count method based on time. The longer a user reads the more words he reads.

### Line-Break Detection

Our approach uses a dynamic line break detection using the distributions of the horizontal component of the saccades. We combine the saccade amplitude $sa$ with the horizontal direction component $sd_h$ of the saccade (-1 for a completely horizontal saccade against main reading direction and +1 for a completely horizontal saccade in reading direction). We refer to it as horizontal saccade direction component (HSD). $HSD = sd_h * sa$. Another way to define HSD is a projection of each saccade on the horizontal axis.

We assume that reading is dominated by two types of saccades, short ones in reading direction (indicating reading words) and longer saccades against the reading direction (indicating the line breaks). Now, if we plot the HSD histogram for some sample reading recording, we should see two maxima (see also Figure 8 for histogram examples), the larger one is the average reading saccade direction * amplitude (forward motion) and the second maximum (smaller as there are fewer line break saccades) in the place of the average line break saccade direction. The text width in coordinates or the so-called line break distance is the distance between theses maxima. Using half of this length as a threshold works well for recognizing a line break (experimentally determined).

Before, we can calculate the HSD histogram on optical eye tracking raw eye gaze data we have to do an additional step,

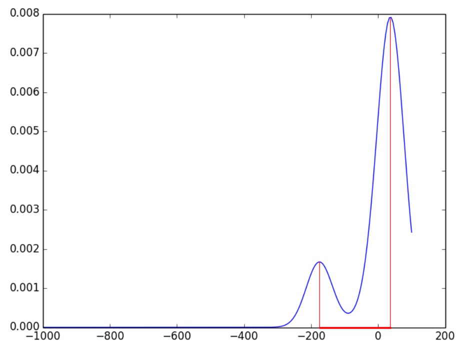



Figure 3: Determining the Line Break Saccade Threshold: Distance calculation between the two largest maxima of a Gaussian mixture fit on the saccade length histogram.

combining consecutive saccades against the reading direction (line break saccades). Sometimes backward saccades of the optical system are separated by small fixations (this might be an artifact of the optical system we use for our recordings). Figure 2 illustrates the effect of this pre-processing step. The top figure shows the eye gaze with 3 backward saccades, the bottom shows the saccades combined into one.

To determine a line break we take all saccades for the segment detected as reading, we calculate the HSDs for each saccade, combine them in the HSD histogram, fit a mixture of 2 Gaussians to it and take the distance between the two maxima (see Figure 3). Empirically, we found that taking half of this distance as threshold for a line break works best.

For our experiments and analysis we assume that the main reading direction is horizontal (right to left or left to right). However, the algorithm is easily adjustable for other reading directions (e.g. for some Japanese/Chinese texts).

### Word Count Estimation

We use two methods to derive the amount of words read.

*Basic Word Count-*
This simple estimator uses the average word count of the document read times the line numbers estimated from the previous step. For this estimator to work we need to get the average words per line for the document. This can be easy achieved if reading on an electronic device. For reading from paper with the optical system, we use a document image retrieval technique called "Locally Likely Arrangement Hashing" (LLAH) to associate the paper with the digital document [39]. We use the video feed from the eye tracker as input to LLAH. LLAH retrieves the corresponding page from a document image database by comparing feature points. This comes with the limitation that the document needs to be registered first in a database, yet as previous research shows the retrieval algorithm is very fast. For example, retrieval from a database with 100 Million pages (around 440 thousand books) takes around 178 ms on a single server core (for performance details please see [30]). Even a corpus as large as Google books can be handled.

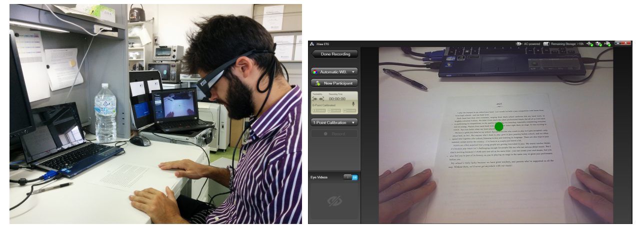


Figure 6: Baseline Experimental Setup. Top picture shows a participant starting to read wearing the mobile eyetracker; the bottom picture shows the recording software with a scene image and the eye fixation depicted as green dot. Bottom shows a photo from the recordings a user reading on a kindle.

*Support Vector Regression Count-*
We calculate features over the complete reading segment recognized by the reading detection method. After evaluating over 25 standard features, we use 5: total time read, sum of all saccades distance, sum of the line break saccade distances, number of line breaks and sum of the reading saccade distances. We train a Support Vector Regression with a Radial Basis Function using these features.

## EXPERIMENTAL SETUPS

### Mobile Optical Eye Tracking

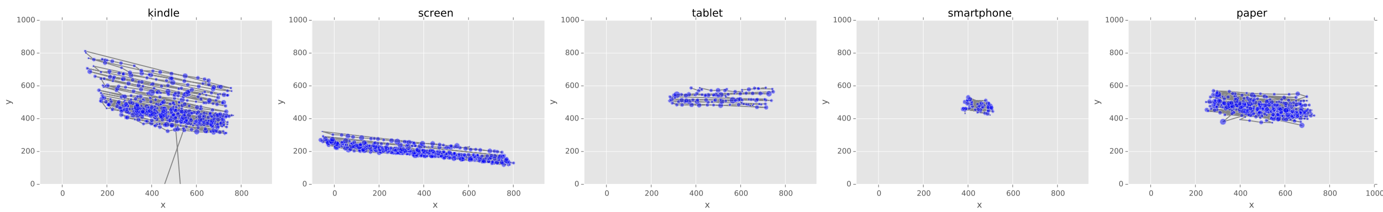
For both optical eye tracking experiments we use the SMI mobile eye tracker 2.0. The glasses have binocular gaze estimates at a joint sampling frequency of 30Hz as well as a scene video with a resolution of 1280x960 pixels. At this sampling frequency, only saccades of about 33ms and slower can be detected. The scene videos are recorded solely for ground truth and documentation purposes. Following the recordings on a laptop, the data is exported by BeGaze 3.4.52, an eye tracking analysis software. The data analysis is done using python scripts, we will make data as well as source code publicly available for other researchers to use.

*Baseline Experiment*
We record the documents used by our previous experiment to extend their validity to international participants with varying English skills [35]. The subject reads a paper in an office scenario. We calibrated the eye tracker using a standard 3-point calibration prior to each recording. Each subject reads 14 documents, the document order is assigned using the Latin Square method.

The documents consist of 10 preliminary English texts (PET) and 4 difficult English Scholastic Assessment Test (SAT) texts. The document size ranges from 135 to 414 words (mean of 245 words). We recruited 9 subjects with an international background with following national background: French, German, Luxembourgian, Beninese, Turkish, Vietnamese (3 female average age 28 std 7). We record the eye gaze and scene camera using the SMI glasses for all participants while reading the documents. The participants don't get any special instructions except of reading naturally. After finishing each document, we ask several comprehension questions to assess the participant in terms of understanding.

(a) (b) (c) (d) (e)

Figure 4: Photos from the Device Experimental Setup with the users reading from different devices: an e-ink reader (a), computer screen (b), tablet computer (c), smartphone (d) and a sheet of paper (e).




(a) (b) (c) (d) (e)

Figure 5: Sample, filtered eye gaze recorded by the SMI mobile eye tracking glasses while the user reads from different devices: an e-ink reader (a), computer screen (b), tablet computer (c), smartphone (d) and a sheet of paper (e).

### *Devices Experiment*

In this experiment we want to evaluate the effects of document length and device types (line length) on the word count accuracy. The subjects read 5 documents with 115, 253, 519, 679 and 881 words respectively. The document are read from different devices. The devices have different width and height sizes, but their font size -12pt- is constant. In the following we list the devices used and their sizes: Paper (169mm width,117mm height), E-ink reader (90mm width, 122mm height), Smartphone (56mm width, 80mm height), Computer Screen (516mm width, 325mm height), Tablet (217mm width, 118mm height). The line breaks differ for each document depending on the device.

We recruited again 10 subjects with an international background (German, Japanese, French, Luxembourgian, Beninese, Australian, Chinese) (4 female, average age 27 std 7). Two subjects have native English skills. They read from different device types (paper, screen, smart phone, tablet, e-ink reader) with varying line lengths and in addition they perform the following activities: solving a Sudoku puzzle on a printed paper, talking to a person, playing Angry Birds on the smart phone, performing a visual search task finding x in the web browser on the computer screen, watching a short movie on the tablet. We again calibrated the eye tracker using a standard 3-point calibration prior to each recording. The document, device and activity assignments are again determined using the Latin Square method. The participants are instructed to perform each task naturally, we don't present them with any specific restrictions regarding the tasks. As with the previous experiment, we ask questions after each document reading task to assess comprehension.

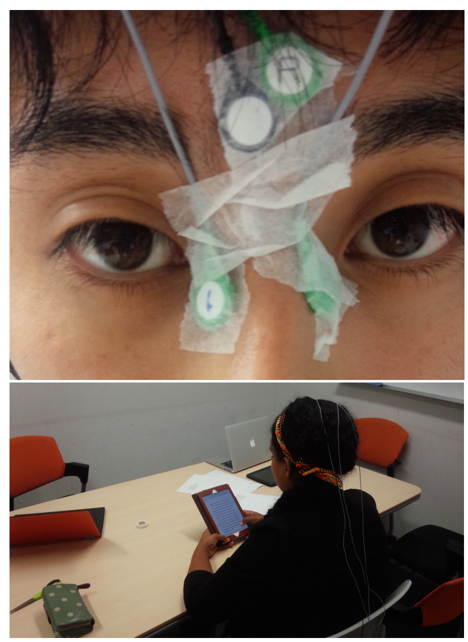



Figure 7: Electrooculography Setup: top shows the electrode placement, we use the potential between electrodes left and right from the nose for the horizontal EOG component and the potential between right/left electrode and top electrode for the vertical EOG compoenent. R is the reference electrode.

## Electrooculography

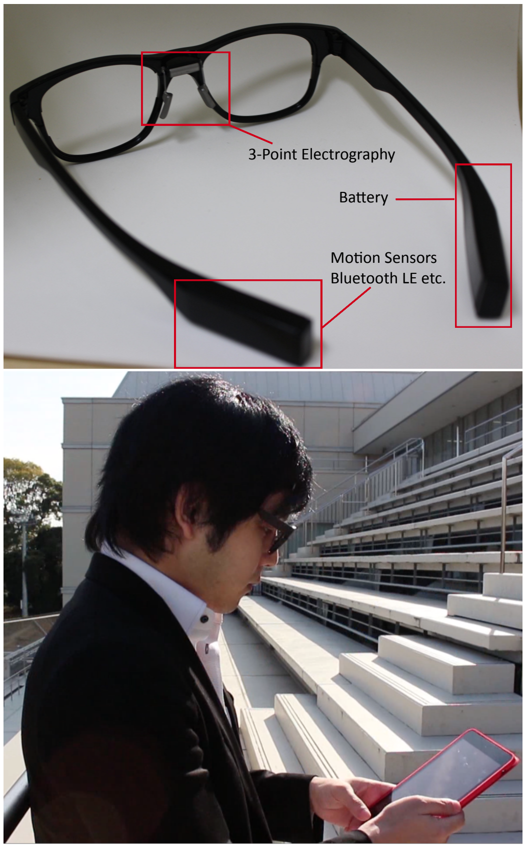



Figure 9: J!NS MEME Prototype and user wearing MEME reading on a iPad outside.

For the EOG experiment, we follow the script of the Devices Experiment given above. We use the data from 6 participants from Kunze et. al. [33] as well as 2 additional users in total 8 (4 female, average age 24 std 8). As EOG Device we use the Polymate mini with active electrodes (sampling rate 1k Hz). The Figure 7 shows our electrode setup. The Polymate is connected via Bluetooth to a laptop. Otherwise the experimental conduction and conditions resemble the previous device experiment.

### MEME Case Study

In addition we use a small case study with an early prototype of commercial smart eyewear, J!NS MEME [1] to show the potential of our approach. J!NS MEME are smart glasses glasses including 3 electrodes to detect eye movements and inertial motion sensors (accelerometer/gyroscope) for head motions (see Figure 9) [1]. MEME streams the sensor data over bluetooth LE to a laptop or smartphone with a sampling rate of over 100 Hz for the EOG data. Battery runtime for the current prototype is around 8 hours. They weight 32 grams and can be easily confused with normal eye glasses.

We record 4 participants (2 female, mean age 34, std. 15) for 3 days (each 2 x 3 hours per day) with MEME. For stability reasons the recording is done on a small laptop (11”).

[1] https://www.jins-jp.com/jinsmeme/en/

The users wear the device. The experiment conductor record when their are reading and the word amount using a custom labeling software on the laptop (start time/end time, word count). The experiment conductor videotapes each run. The users are free to choose texts, reading devices, location and other activities recorded. We have however 2 limitations: (1) the material read must consist mostly of text (no comics etc.) (2) for the free activities we exclude text processing (writing text on paper or computer). We believe we can deal with more free reading material and text processing to some extend, yet it would complicate the word counting (especially ground truth recordings) and is left for future work. Activities done by users include: cooking a meal, making coffee, photocopying, ironing clothes. Reading done by users include: novel on iPad outside, newspaper in coffeeshop, textbook in car etc.

## RESULTS AND DISCUSSION

In the following, we present the analysis results for the different data sets. All evaluations are done using the leave-one-out strategy, training on n-1 users, evaluating on 1 user this n times, presenting the average over all as results.

The reading detection performs very well (close to 100 % for all data sets except the MEME Use Case see later). Therefore, we will not discuss it further.

### Optical Eye Tracking

Figure 8 depicts the HDS histograms for the different reading devices of the device experiment. Our line break detection method decreases the error from 15% to 6% for the Baseline Dataset and from 62% to 8% for the Devices Dataset comparing it with the algorithm of the previous work.

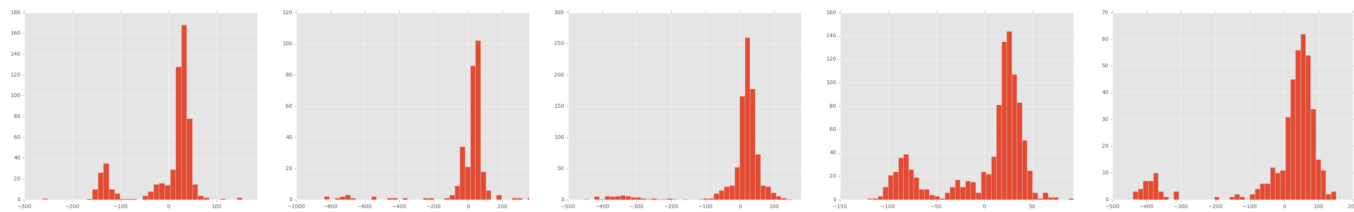
### Word Count Estimation

| Method | Baseline Exp | Devices Exp |
|---|---|---|
| Time | 22% | 32% |
| Previous Method[35] | 11% | 45% |
| Basic Word Count (average words * lines) | 9% | 8% |
| Support Vector Regression Word Count | 8% | 17% |

Table 2: Overview about the word count estimation error for different methods

Table 3 summarizes the results for estimating the words read for the different data sets. We see that the most basic estimation using Reading detection and Time has an error of 22 % or 32%. The previous method by Kunze et al. performs better in the Baseline Experiment, yet worse in the Devices Experiment, as it cannot cope well with changes in line length. The Basic Word Count using average words per line performs best overall with 9% and 8%. Support Vector Regression has trouble with the Devices Experiment with a 17% error rate. This is mostly due to the very diverse English skills of the subjects participating in this experiment.

The standard deviation of the error rate is 3% for baseline 4% for devices experiment. Compared to a std of 15 % and 35% of the previous method and 22% and 45% for using time only. The improvement is also significant shown by an F-Test between the error rates of the Kunze et al. method and the current implementation ($p<0.05$, $F(4,6)= 5.14$).



(a) (b) (c) (d) (e)

Figure 8: Horizontal Saccade Direction (HSD) Histograms for different reading devices: an e-ink reader (a), computer screen (b), tablet computer (c), smartphone (d) and a sheet of paper (e).

Reading on the Computer Screen has by far the highest error. This is reasonable, as there are very few, long lines (missing one line or detecting an additional line increases the error significantly) and the reading is very unnatural. Users tend to move their head a lot while reading. The second worst device is the smartphone, as in this case the users also moved a lot their head and the device.

### Electrooculography

For the EOG, we compare our inference to a word count estimate derived from a perfect reading detection system for baseline (most research in the related work focuses on it). We estimate the number of words a person was reading just based on time and compare this to our system.

| Method | Error Rate | STD |
|---|---|---|
| Time Baseline | 31% | 9% |
| Static Word Count | 13% | 3% |
| SVR Word Count | 5% | 0.2% |

Table 3: Overview of the word count estimation error and standard deviation of the error for different methods. First the baseline just using the time a user read a text, second is a static word number times the detected line-breaks, third a SVR based on Line Break features alone, and last a SVR based on Line Break and Line features.

For the Line break detection we have an error of 5%, std 1.2%. The summary of the results can be found in Table3. The static word count method already performs with around half the error of the time baseline (13%).

Comparing optical tracking with EOG, it the EOG system performs better than the mobile eye tracker. Yet, this is expected. Our word count method focuses on line break detection saccades and the EOG is usually preferred for saccade detection as it can sample higher, in our case 1k Hz compared to 60 Hz for the SMI eye tracker. Additionally, the advantage of the optical system (gaze coordinates not only eye movement data) are not relevant for our application case.

### MEME Case Study

For reading detection, recall is 87% precision is only at 79 %, yet looking into the data, one particular prototype shows a very noisy EOG signal (additional frequencies around 20 - 50 Hz). The errors happen exactly with this prototype for 3 hours of the total recordings. Removing these 3 hours we get to 89% precision. The SVR word count is at 20 % error rate comparing it to 30 % on the data using perfect reading detection and time only.

### Improvements and Limitations

A major problem not addressed in previous work are changing line lengths. Therefore we see the dynamic line-break detection as one of our major contributions for this work, applicable to wide variety of eye movement data. We got also rid of the more direction criterion introduced by Kunze et. al (the direction of the saccade must be slightly downwards). This makes the line break detection is robust against "swiping while reading" e.g. on a tablet. Rereading some words can be also tolerated to a certain extend (as long as it is only half the width). However, if there are short lines (again smaller than half the width of a regular line), our algorithm fails to detect them. This trade-off can be adjusted by the line break threshold, if we decrease it smaller to half the size of a normal line break saccade, the effect of re-reading on the word count is stronger yet shorter lines can still be detected.

Given the respective 17% or 8% of total error from our experiment, we try to assess the question: Is our method accurate enough to quantify reading habits in a sensible way for users? This is difficult to address without a system implementation and long term study. Yet, when we compare the performance to physical activity tracking we can find an answer. We assume that the tracking accuracy for inducing physical and mental behavior changes are equivalent. In most controlled experimental setups, step counters show an error rate of 5-10%, the error can increase rapidly in real-life scenarios to up to 20 - 35%[47]. Still they are regarded as a effective measure to log activity and motivate people to become more active. Therefore, we believe the perceived error of 8% to 17% is enough to motivate users to read more. Comparing our method to step counting might be controversial, as we are dealing with a cognitive task and not physical activity. However, studies in learning and reading have shown that learners who track their work manually feel encouraged and acquire faster and better reading skills [27].

Of course, larger scale studies are needed to validate the accuracy of our method for more real-life datasets (e.g. including reading in different situations, not only an office environment). Also, long term studies have to show if cognitive activity tracking is as motivating as physical activity tracking given the same level of accuracy.

Currently, the word count estimation is done in batch processing on previously recorded eye gaze data. In principle, all algorithm steps work online (given the 3 second delay for the reading detection). The dynamic line break detection is more problematic, as it requires a couple of line breaks to work and with fewer line breaks the results might not be robust enough. This should be investigated in future work. The computational complexity of the methods is quite low. The hardest is the SVR training phase, gaze filtering, line break detection and classification can be done in polynomial time. Taking the line break detection into account, a robust word count estimation could be done in a least minute interval maybe even shorter.

## CONCLUSIONS AND FUTURE WORK

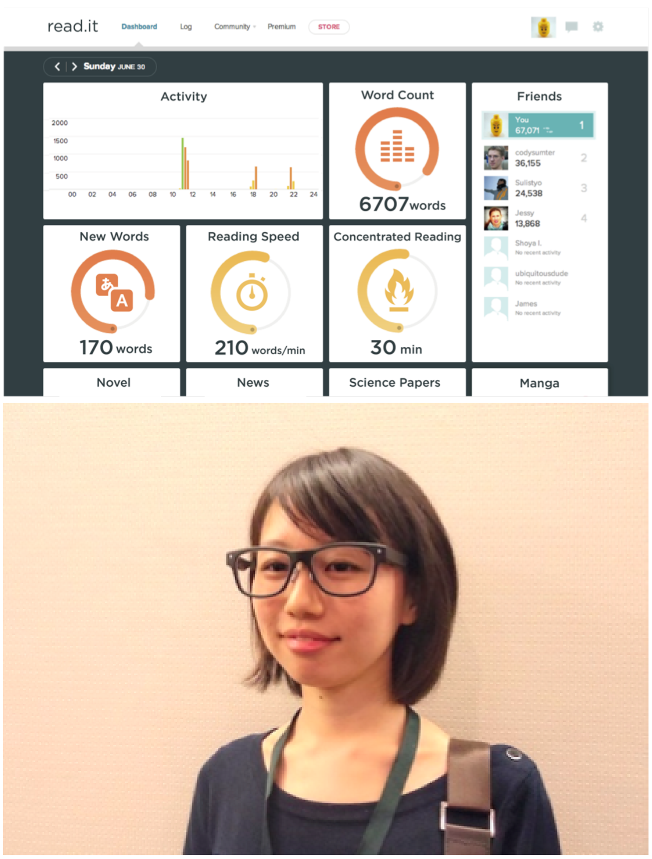



Figure 10: Mockup of a reading service using our approach and a user wearing J!NS MEME prototype, smart glasses with EOG.

We presented our work towards tracking how much a user reads, enabling quantified feedback about reading volume. We show an word count estimation algorithm that works with 8% (using some kind of document identification) or 17% just on eye gaze only, user-independent evaluated on 2 experiments with a total of 19 users. In addition we evaluate our method also on an EOG recording.

*Larger Scale Experiments-*
An obvious extension to our current work is to include more participants, increase the device diversity and also change font sizes. We expect that the dynamic line break detection algorithm should also work in these situations. Yet, especially we need to record more data with native speakers as the results show the estimates for them are not so good. This indicates that reading skill level impacts the word estimation method strongly. This raises the question if we can use this to our advantage and estimate the reading skills using eye gaze features.

*General Cognitive Activity Tracking using Visual Behavior-*
As mentioned, just the saccade features used for reading detection are enough to separate the different other activities we recorded (e.g. watching video, playing games, ...) reasonably well (71%). We should explore how far visual behavior can be used to distinguish and detect various cognitive activities.

*Extending over Document Types-*
So far we only use texts from reading comprehension sections of standardized tests. This gives us so far the ability to assess the reading skill of the user (and ensure the user actually read the text and is not only faking reading). Other document types could also be recognized using information about text layout, this in turn can help to improve the word count estimation.

*Word Count Estimation for Everybody-*
To enable word count estimation for a larger population, we need to port the system to one of the hopefully soon available commercial EOG devices or cheap DIY eye trackers. Smart eye glasses seem to be the best platform to achieve this goal [1]. We plan on implementing our system on J!NS MEME (see Figure 10). We expect that the algorithm can work with slight adjustments, given a good EOG signal quality.

## ACKNOWLEDGEMENTS

This work is supported in part by the JST CREST project ”Reading Life Log”. Part of this work was envisioned at the Dagstuhl Seminar ”Augmenting Human Memory” No. 14362.

## REFERENCES

1. Amft, O., Wahl, F., Ishimaru, S., and Kunze, K. Making regular eyeglasses smart. *Pervasive Computing, IEEE 14*, 3 (2015), 32–43.

2. Biedert, R., Buscher, G., and Dengel, A. The eyebook–using eye tracking to enhance the reading experience. *Informatik-Spektrum 33*, 3 (2010), 272–281.

3. Biedert, R., Buscher, G., Schwarz, S., Hees, J., and Dengel, A. Text 2.0. In *Ext. Abs. CHI 2010* (2010), 4003–4008.

4. Biedert, R., Dengel, A., Elshamy, M., and Buscher, G. Towards robust gaze-based objective quality measures for text. In *Proc. ETRA 2012* (2012), 201–204.

5. Biedert, R., Hees, J., Dengel, A., and Buscher, G. A robust realtime reading-skimming classifier. In *Proc. ETRA 2012* (2012), 123–130.

6. Bravata, D. M., and Smith-Spangler. Using pedometers to increase physical activity and improve health. *The journal of the American Medical Association 298*, 19 (2007), 2296–2304.

7. Bulling, A., and Roggen, D. Recognition of visual memory recall processes using eye movement analysis. In *Proceedings of the 13th international conference on Ubiquitous computing*, ACM (2011), 455–464.

8. Bulling, A., Ward, J., Gellersen, H., and Troster, G. Eye movement analysis for activity recognition using electrooculography. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 33*, 4 (2011), 741–753.

9. Bulling, A., Ward, J. A., and Gellersen, H. Multimodal Recognition of Reading Activity in Transit Using Body-Worn Sensors. *ACM Trans. on Applied Perception 9*, 1 (2012), 2:1–2:21.

10. Bulling, A., Ward, J. A., Gellersen, H., and Tröster, G. Eye Movement Analysis for Activity Recognition Using Electrooculography. *IEEE Trans. on Pattern Analysis and Machine Intelligence 33*, 4 (Apr. 2011), 741–753.

11. Bulling, A., and Zander, T. O. Cognition-aware computing. *Pervasive Computing, IEEE 13*, 3 (2014), 80–83.

12. Buscher, G., and Dengel, A. Gaze-based filtering of relevant document segments. In *WSRSP Workshop*, vol. 9 (2009), 20–24.

13. Buscher, G., Dengel, A., and van Elst, L. Eye movements as implicit relevance feedback. In *Ext. Abs. CHI 2008* (2008), 2991–2996.

14. Buscher, G., Dengel, A., van Elst, L., and Mittag, F. Generating and using gaze-based document annotations. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '08, ACM (New York, NY, USA, 2008), 3045–3050.

15. Carver, R. P. *Sense and nonsense in speed reading*. Revrac Publications, 1971.

16. Clarke, A. R., Barry, R. J., McCarthy, R., and Selikowitz, M. Eeg analysis of children with attention-deficit/hyperactivity disorder and comorbid reading disabilities. *Journal of Learning Disabilities* (2002), 276–285.

17. Cunningham, A., and Stanovich, K. What reading does for the mind. *Journal of Direct Instruction 1*, 2 (2001), 137–149.

18. Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A., and Kliegl, R. Coregistration of eye movements and eeg in natural reading: analyses and review. *Journal of Experimental Psychology: General 140*, 4 (2011), 552.

19. Ehrlich, E. Opinions differ on speed reading. *NEA Journal 52* (1963), 43–44.

20. Ferstl, E. C., Neumann, J., Bogler, C., and von Cramon, D. Y. The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping* (2008), 581–593.

21. Genesee, F., Lindholm-Leary, K., Saunders, W., and Christian, D. English language learners in us schools: An overview of research findings. *Journal of Education for Students Placed at Risk 10*, 4 (2005), 363–385.

22. Gevins, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., and Rush, G. Monitoring working memory load during computer-based tasks with eeg pattern recognition methods. *Human Factors: The Journal of the Human Factors and Ergonomics Society 40*, 1 (1998), 79–91.

23. Gog, T. v., Kester, L., Nievelstein, F., Giesbers, B., and Paas, F. Uncovering cognitive processes: Different techniques that can contribute to cognitive load research and instruction. *Computers in Human Behavior 25*, 2 (2009), 325–331.

24. Graf, R. G. Speed reading: Remember the tortoise. *Psychology Today 7*, 7 (1973), 112–13.

25. Halpern, D. F. Teaching for critical thinking: Helping college students develop the skills and dispositions of a critical thinker. *New directions for teaching and learning 1999*, 80 (1999), 69–74.

26. Hansen, D. M. A discourse structure analysis of the comprehension of rapid readers.

27. Hattie, J. *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge, 2013.

28. Iqbal, S. T., Zheng, X. S., and Bailey, B. P. Task-evoked pupillary response to mental workload in human-computer interaction. In *CHI'04 extended abstracts on Human factors in computing systems*, ACM (2004), 1477–1480.

29. Irwin, D. E. Fixation location and fixation duration as indices of cognitive processing. *The interface of language, vision, and action: Eye movements and the visual world* (2004), 105–134.

30. Iwamura, M., Nakai, T., and Kise, K. Improvement of retrieval speed and required amount of memory for geometric hashing by combining local invariants. *Proc. of BMVC2007* (2007), 1010–1019.

31. Just, M. A., and Carpenter, P. A. *The psychology of reading and language comprehension.* Allyn & Bacon, 1987.

32. Kliegl, R., Nuthmann, A., and Engbert, R. Tracking the mind during reading: the influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General; Journal of Experimental Psychology: General 135*, 1 (2006), 12.

33. Kunze, K., Katsutoshi, M., Uema, Y., and Inami, M. How much do you read? counting the number of words a user reads using electrooculography. In *Proceedings of*

*the 5th Augmented Human International Conference*, ACM (2015), 150–153.

34. Kunze, K., Kawaichi, H., Yoshimura, K., and Kise, K. Towards inferring language expertise using eye tracking. In *Ext. Abs. CHI 2013* (2013), 4015–4021.
35. Kunze, K., Kawaichi, H., Yoshimura, K., and Kise, K. The wordmeter – estimating the number of words read using document image retrieval and mobile eye tracking. In *Proc. ICDAR 2013* (2013).
36. Kunze, K., Yuki, S., Ishimaru, S., and Kise, K. Reading activity recognition using an off-the-shelf eeg. In *Proc. ICDAR 2013* (2013).
37. Martínez-Gómez, P., and Aizawa, A. Recognition of understanding level and language skill using measurements of reading behavior. In *Proceedings of the 19th international conference on Intelligent User Interfaces*, ACM (2014), 95–104.
38. Moore, D. W., Alvermann, D. E., and Hinchman, K. A. *Struggling Adolescent Readers: A Collection of Teaching Strategies.* ERIC, 2000.
39. Nakai, T., Kise, K., and Iwamura, M. Use of affine invariants in locally likely arrangement hashing for camera-based document image retrieval. *In Proc. of DAS 2006 3872* (Feb. 2006), 541–552.
40. Orchard, L. N., and Stern, J. A. Blinks as an index of cognitive activity during reading. *Integrative Physiological and Behavioral Science 26*, 2 (1991), 108–116.
41. Patterson, K. E., Graham, N., and Hodges, J. R. Reading in dementia of the alzheimer type: A preserved ability? *Neuropsychology 8*, 3 (1994), 395.
42. Rashotte, C. A., and Torgesen, J. K. Repeated reading and reading fluency in learning disabled children. *Reading Research Quarterly* (1985), 180–188.
43. Rayner, K. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin 124*, 3 (1998), 372.
44. Rudmann, D. S., McConkie, G. W., and Zheng, X. S. Eyetracking in cognitive state detection for hci. In *Proceedings of the 5th International Conference on Multimodal Interfaces*, ICMI '03, ACM (New York, NY, USA, 2003), 159–163.
45. Spache, G. D. Is this a breakthrough in reading? *The Reading Teacher 15*, 4 (1962), 258–263.
46. Strait, M., Canning, C., and Scheutz, M. Reliability of nirs-based bcis: a placebo-controlled replication and reanalysis of brainput. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, ACM (2014), 619–630.
47. Swartz, A., Bassett Jr, D., Moore, J., Thompson, D., and Strath, S. Effects of body mass index on the accuracy of an electronic pedometer. *International journal of sports medicine 24*, 08 (2003), 588–592.
48. Tsunashima, H., Yanagisawa, K., and Iwadate, M. Measurement of brain function using near-infrared spectroscopy (nirs), 2012.
49. Xu, S., Jiang, H., and Lau, F. C. User-oriented document summarization through vision-based eye-tracking. In *Proc of IUI*, IUI '09 (2009), 7–16.