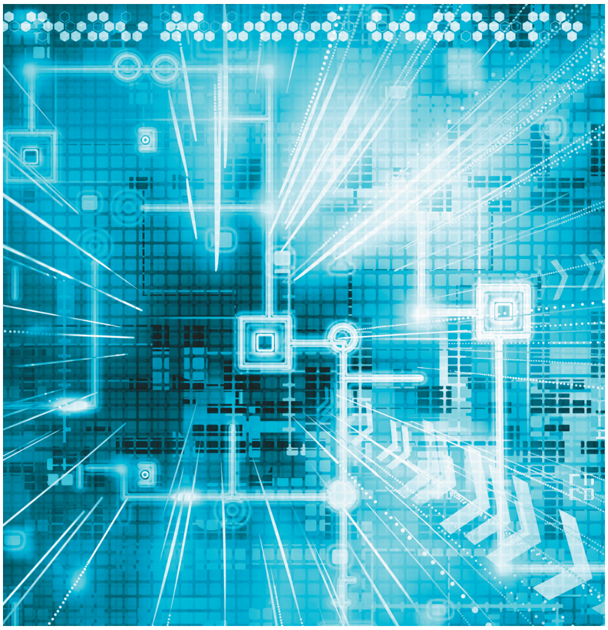

Activity Recognition for the Mind: Toward a Cognitive "Quantified Self"

Kai Kunze, Masakazu Iwamura, and Koichi Kise, *Osaka Prefecture University*

Seiichi Uchida, *Kyushu University*

Shinichiro Omachi, *Tohoku University*

Applying mobile sensing technology to cognitive tasks will enable novel forms of activity recognition.

Physical activity recognition technology has become mainstream—many dedicated mobile devices and smartphone apps count the steps we climb or the miles we run. What if devices and apps were also available that could count the words we read and how far we've progressed in our learning? The authors of this article demonstrate that mobile eye tracking can be used to do just that. Focusing on reading habits, they've prototyped cognitive activity recognition systems that monitor what and how much users read as well how much they understand. Such systems could revolutionize teaching, learning, and assessment both inside and outside the classroom. Further, as sensing technology improves, activity recognition could be extended to other cognitive tasks including concentrating, retaining information, and auditory or visual processing. While this research is extremely exciting, it also raises numerous ethical questions—for example, who should know what we read or how much we understand?

Albrecht Schmidt, column editor

People increasingly use mobile computing technology to track their health and fitness progress, from simple step counting to monitoring food intake to measuring how long and well they sleep. Smartphone applications such as RunKeeper (http://runkeeper.com) and Lose It! (www.loseit.com) and wearable devices such as the Fitbit FLEX wristband (www.fitbit.com) foster better eating and exercise habits, decrease the risk of obesity-related diseases, and improve quality of life.

Activity-tracking abilities are still hampered by the limited battery power of today's mobile devices, but emerging technologies such as the M7 motion-sensing coprocessor in the new iPhone 5s make it easier to aggregate and interpret sensor data in a power-efficient manner. In addition, while most activity-recognition apps continue to focus on physical movement, such as steps taken or stairs climbed, new products such as Withings Pulse (www.withings.com/en/pulse) also quantify physiological signals such as heart rate.

Given these trends, it's only a matter of time before we see mobile sensing technology applied to cognitive tasks, enabling novel forms of activity recognition.

EYE MOVEMENTS AND COGNITIVE TASKS

Sensing brain activity seems to provide the most insight into cognitive processes. However, doing so accurately requires relatively invasive methods such as electroencephalography (EEG), functional magnetic resonance imaging, and electrocorticography. Of these, EEG is the most promising for mobile applications, as evidenced by forthcoming products such as the Emotiv Insight neuroheadset (http://emotivinsight.com). Yet, EEG is quite noisy and easily overshadowed by

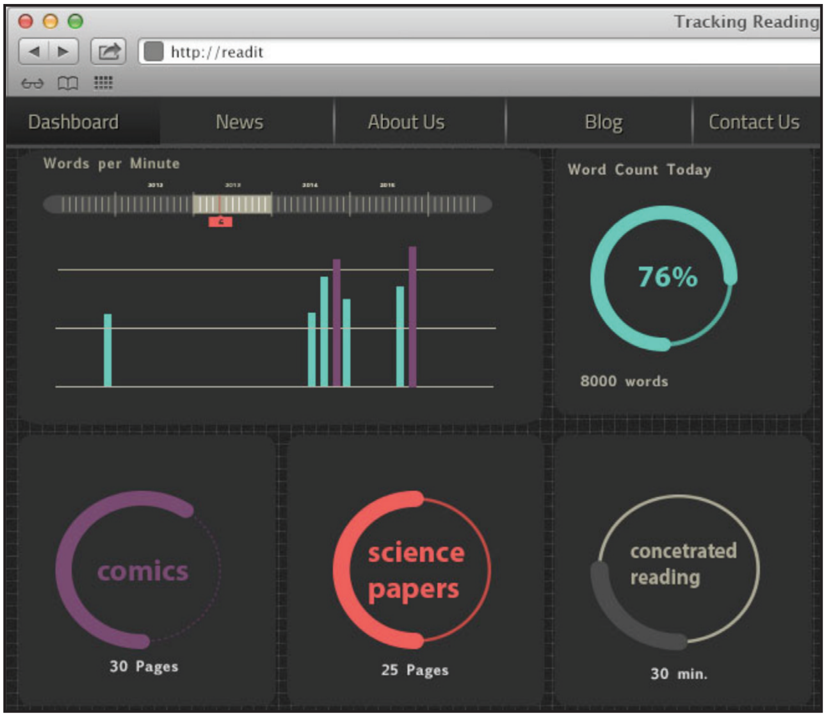

Figure 1. Mockup interface for a hypothetical Readit service that records and analyzes reading habits, much like a fitness service such as Fitbit.

muscle movement, requiring time-consuming and difficult signal processing.

An alternative approach is to track eye movements, which are strongly correlated with cognitive tasks. However, because eye movements include information about the user's attentiveness, degree of fatigue, emotional state, and so on, it can be difficult to isolate the object of interest. Nevertheless, early research showed that stationary eye trackers can differentiate various types of media consumption—for example, watching videos, surfing the Web, and reading magazines have different eye-movement characteristics (H. Drewes and A. Schmidt, "Interacting with the Computer Using Gaze Gestures," *Proc. 11th IFIP TC Int'l Conf. Human-Computer Interaction* [Interact 07], LNCS 4463, Springer, 2007, pp. 475-488).

MOBILE EYE TRACKING

Electrooculography (EOG) and optical eye tracking are two promising methods to record eye gaze in real-life mobile settings.

The eye can be represented as a dipole between the cornea and retina, and EOG uses electrodes to measure the change in potential when the eye moves. With this technique, it's possible to recognize reading and other eye movements in realistic everyday scenarios. Although the electrodes to measure EOG can be easily embedded in glass frames to make them more unobtrusive, they still need to touch the skin near the eye, which can be uncomfortable. On the other hand, this approach is cheap to implement and requires little processing.

An alternative option is to track eye gaze using infrared light and stereo cameras, which provides potentially higher accuracy but requires more processing power. This type of system infers eye motion and gaze pointing based on iris shape, sometimes in combination with local face features such as eyelids and eye corners.

Commercial mobile eye trackers are still quite expensive. Yet, given that the hardware—electrodes, cameras, and infrared LEDs—is relatively inexpensive, prices should fall as demand increases just as they have for head-mounted displays since Google Glass was announced. An example of a cheap 3D-printed eye tracker was shown at this year's MobileHCI conference (K. Lukander et al., "OMG!: A New Robust, Wearable and Affordable Open Source Mobile Gaze Tracker," *Proc. 15th Int'l Conf. Human-Computer Interaction with Mobile Devices and Services*, ACM, 2013, pp. 408-411).

Patent applications suggest that Google, Apple, and other tech companies are already trying out mobile eye-tracking prototypes internally. Researchers are also experimenting with front cameras on tablets and smartphones to carry out low cost, low-fidelity eye-gaze detection (K. Kunze et al., "My Reading Life: Towards Utilizing Eyetracking on Unmodified Tablets and Phones," *Proc. UbiComp '13 Adjunct*, ACM, 2013, pp. 283-286).

QUANTIFYING READING HABITS

Measuring reading activity is one way to quantify mental "fitness." Although an increase in reading volume leads to demonstrably improved language comprehension and critical thinking skills, few eye-tracking researchers have evaluated reading habits in situ (A. Bulling, J.A. Ward, and H. Gellersen, "Multimodal Recognition of Reading Activity in Transit Using Body-Worn Sensors," *ACM Trans. Applied Perception*, vol. 9, no. 1, 2012, article no. 2).

In our research, we're exploring the possibility of using mobile eye-tracking technology to determine how much people read, what type of document they're reading, and how much of what they're reading that they understand.

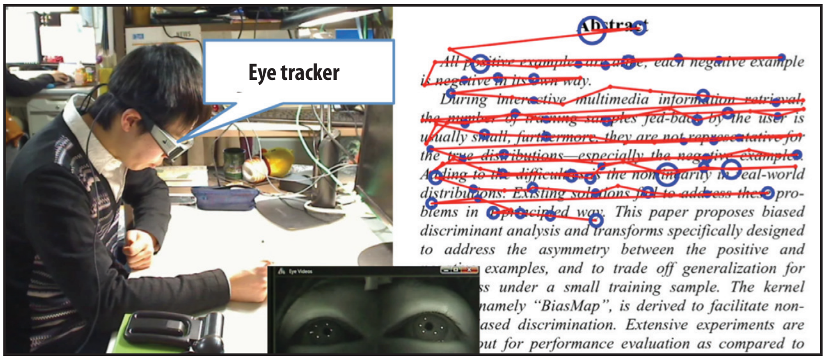

Figure 2. Using mobile eye tracking to count read words. Left: A user reads while wearing the SMI iView X eye tracker. Lower center: The eye tracker uses infrared light and stereo camera views to track eye gaze. Right: The Wordometer application maps the user's eye gaze to the document he's reading. The lines show saccades, and the circles indicate where users fixated their gaze. The size of the circles correlates with the duration of the fixations.

Just as fitness apps record and analyze physical activity to help users get into better shape, a "reading habits" app could record and analyze the number of words read, types of documents read, and periods of concentrated reading to help users become better thinkers. Figure 1 shows an interface mockup for a hypothetical Readit service.

Counting read words

Toward this end, we've created the Wordometer app for the SMI iView X eye tracker, shown in Figure 2, that estimates the number of words a person reads, much like a pedometer counts steps (K. Kunze et al., "The Wordometer—Estimating the Number of Words Read Using Document Image Retrieval and Mobile Eye Tracking," *Proc. 12th Int'l Conf. Document Analysis and Recognition* [ICDAR 13], IEEE CS, 2013, http://kaikunze.de/papers/2013Kunze-1.pdf).

To develop the app, we first recorded the eye movements of numerous users reading English documents and then applied a line-break detector to the data that looked for a long saccade (fast eye movement) against the main left-to-right reading direction. Finally, we created a support vector regression algorithm that inputs detected line breaks and some eye-gaze features to estimate how many words a user has read. Our word-count estimation method achieves an average error rate of only 6.5 percent for 8 users over 10 documents.

Distinguishing document types

Counting read words is just one aspect of understanding reading habits. We're also using mobile eye tracking to distinguish different document types (K. Kunze et al., "I Know What You Are Reading—Recognition of Document Types Using Mobile Eye Tracking," *Proc. 17th Ann. Int'l Symp. Wearable Computers* [ISWC 13], ACM, 2013, pp. 113-116; http://kaikunze.de/papers/2013Kunze-5.pdf).

In particular, we analyze saccade directions to obtain information about how much structured text versus images are on a page. The slope of the eye gaze indicates general reading direction, and a percentile distance gives an approximation of page size.

To assess this technique's effectiveness, we conducted a study in which eight users wearing the iView X eye tracker in five different environments read five different types of Japanese reading material (novel, manga, magazine, newspaper, and textbook), spending about 10 minutes on each document. This approach achieved a 99 percent user-dependent and 74 percent user-independent recognition rate over a window size of about 1 minute of gaze data. This research represents a crucial first step toward the vision of wearable reading assistants and reading logs.

Assessing reading comprehension

In addition, we're exploring whether eye tracking can be used to measure reading comprehension—specifically, second-language skills (K. Kunze et al., "Towards Inferring Language Expertise Using Eye Tracking," *CHI '13 Extended Abstracts on Human Factors in Computing Systems* [CHI EA 13], ACM, 2013, pp. 217-222).

We had 10 native Japanese speakers learning English as a second language read 10 textual passages and answer the accompanying comprehension questions from the Test of English for International Communication (TOEIC) while wearing the iView X eye tracker. The participants sat at a regular desk with the test in front of them and had about 20 minutes to complete each task, after which we asked them to mark those words in the passage they didn't understand.

We recorded the subjects' eye gaze and mapped this data to digitized versions of the textual passages using an image-retrieval method based on locally likely arrangement hashing (LLAH), as Figure 3 shows. Figure 4a depicts a raw eye trace recorded by the eye tracker laid over the digital document using LLAH and manual line segmentation. Horizontally projecting the read line results in the fixation points shown in Figure 4b. This makes it possible to calculate a histogram of the eye fixations that

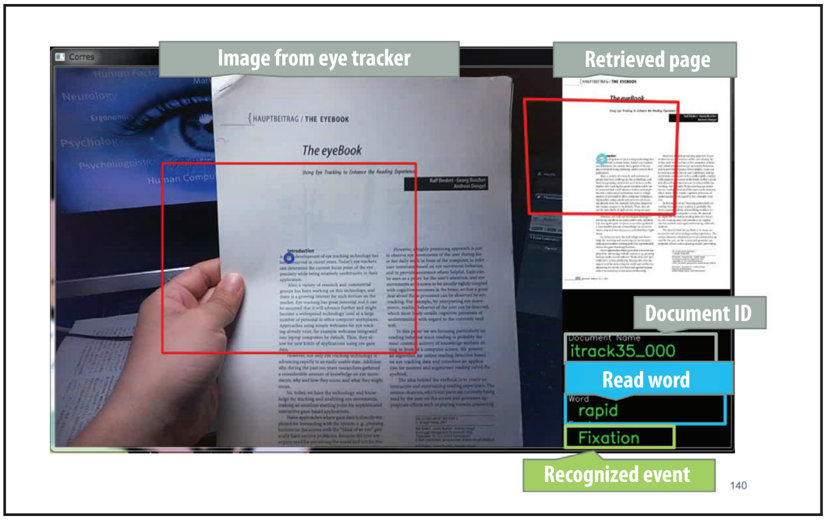

Figure 3. Mapping reading activity recorded using the eye tracker to a digitized version of the physical document retrieved from a database.

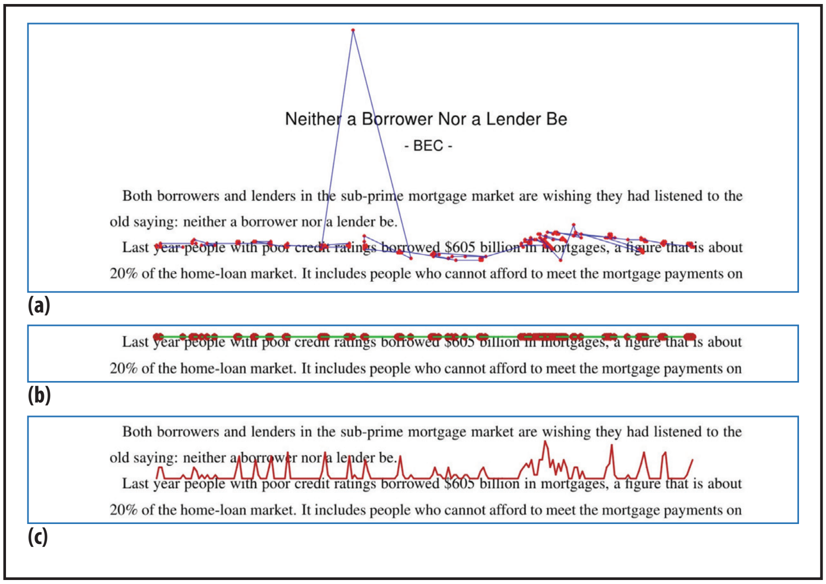

Figure 4. Visualization of eye-movement on a single sentence based on the recording with the eye tracker. (a) Raw trace of eye movements laid over the digital document; the red dots indicate fixation points and the blue lines indicate saccades. (b) Processed eye-movement data: the fixations are filtered and aligned to the text line to provide insights into reading speed and time spend on words. (c) Histogram showing aggregated time spent reading parts of the text; in this case, the user didn't understand the work "mortgages" well.

reveals what words the reader had trouble with, as Figure 4c shows. In this case, the reader had special difficulty understanding "mortgages."

The method worked for all words marked as difficult by the participants. Future research will study how to use eye gaze to infer readers' TOEIC score.

Although our research has focused on reading, the methodology we've developed can be applied to a wide variety of cognitive tasks. Our overall goal is to facilitate long-term tracking of all types of mental processes, which will enable new forms of self-reflection and suggest strategies to optimize mental fitness and well-being.

If educators could obtain feedback about their students' attention level in class and learning progress, for example, they could modify their presentation and tailor learning material to individual pupils' needs, backgrounds, and preferences. Content creators could likewise leverage quantified "mind logs" to improve their works: Which part of a movie most excites users? What feelings does a particular paragraph in a book convey? At what point in a grant proposal does the reader lose interest? And, given enough users and high-quality sensor data, researchers could even address more complex lifestyle issues, such as how sleep and eating habits influence alertness and learning. ■

Kai Kunze *is an assistant professor in the Department of Computer Science and Intelligent Systems at Osaka Prefecture University, Japan. Contact him at kai.kunze@gmail.com.*

Masakazu Iwamura *is an associate professor in the Department of Computer Science and Intelligent Systems at Osaka Prefecture University. Contact him at masa@cs.osakafu-u.ac.jp.*

Koichi Kise *is a professor in the Department of Computer Science and Intelligent Systems at Osaka Prefecture University. Contact him at kise@cs.osakafu-u.ac.jp.*

Seiichi Uchida *is a professor in the Department of Advanced Information Technology at Kyushu University, Japan. Contact him at uchida@ait.kyushu-u.ac.jp.*

Shinichiro Omachi *is a professor in the Department of Communications Engineering at Tohoku University, Japan. Contact him at machi@ecei.tohoku.ac.jp.*

Editor: Albrecht Schmidt, University of Stuttgart, Germany; Albrecht@computer.org