SwitchBack: Using Focus and Saccade Tracking to Guide Users' Attention for Mobile Task Resumption

Alex Mariakakis[1], Mayank Goel[1], Md. Tanvir Islam Aumi[1], Shwetak N. Patel[1], Jacob O. Wobbrock[2]

[1]Computer Science & Engineering | DUB Group
University of Washington
Seattle, WA 98195 USA
{atm15, mayank, tanvir, shwetak}@cs.washington.edu

[2]The Information School | DUB Group
University of Washington
Seattle, WA 98195 USA
wobbrock@uw.edu

ABSTRACT
Smartphones and tablets are often used in dynamic environments that force users to break focus and attend to their surroundings, creating a form of "situational impairment." Current mobile devices have no ability to sense when users divert or restore their attention, let alone provide support for resuming tasks. We therefore introduce *SwitchBack*, a system that allows mobile device users to resume tasks more efficiently. SwitchBack is built upon *Focus and Saccade Tracking (FAST)*, which uses the front-facing camera to determine when the user is looking and how their eyes are moving across the screen. In a controlled study, we found that FAST can identify how many lines the user has read in a body of text within a mean absolute percent error of just 3.9%. We then tested SwitchBack in a dual focus-of-attention task, finding that SwitchBack improved average reading speed by 7.7% in the presence of distractions.

Author Keywords
Situational impairments; mobile; gaze-tracking; reading.

ACM Classification Keywords
H.5.2. Information interfaces and presentation: User Interfaces – *input devices and strategies.*

INTRODUCTION
Mobile computing devices such as smartphones and tablets are now some of the most common devices in our environment. The ubiquity of mobile devices allows users to consume information everywhere. Although these devices are being used in a variety of environments, they do not have any significant awareness of the differences among environments or how these differences affect their users' behaviors. When users are negatively impacted in their ability to interact with technology by environmental conditions, they can be said to be *situationally impaired* [31]. Situational impairments may be caused by a number of factors, such as motion, temperature, ambient noise, *etc*.

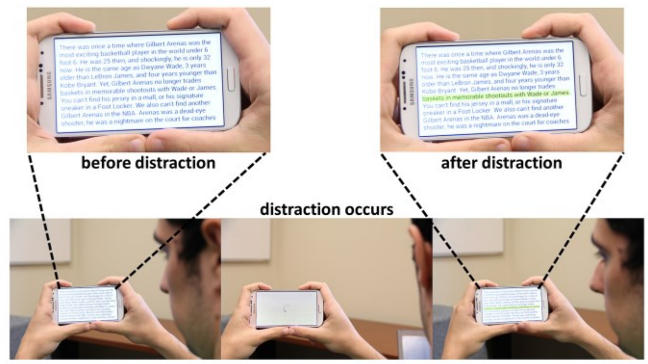

Figure 1. SwitchBack highlights where the user was last looking in a body of text before he or she turned away to handle a distraction.

One of the most significant contextual factors that affects people's mobile device usage is divided attention. For example, if pedestrians are checking email on their smartphones while walking across the street, they must break their attention from their devices to maintain awareness of their surroundings, or else they put themselves at risk of physical injury. Studies have shown that mobile phone related injuries among pedestrians doubled between 2005 and 2010, reaching almost 4% of all pedestrian accidents [27]. Some cities have even begun issuing tickets to pedestrians who are caught texting while walking [28]. Although safety is of the utmost concern when it comes to situational impairments, damage to users' productivity is also a concern. In the aforementioned scenario, it is likely that when pedestrians return their attention to their mobile devices, they will have lost track of their progress. Context-switching incurs a startup cost that can accumulate to the point where users' comprehension is negatively affected [25].

To address these concerns, we present *SwitchBack* (Figure 1), a system that uses the camera on a mobile device to determine when the user is unable to pay visual attention to the device, pause the task (if applicable), and then help the user to efficiently resume the task when the user returns his or her gaze to the device. SwitchBack is built upon our

CHI 2015, April 18 - 23 2015, Seoul, Republic of Korea
Copyright 2015 ACM 978-1-4503-3145-6/15/04…$15.00
http://dx.doi.org/10.1145/2702123.2702539

underlying camera-based attention-tracking algorithm called *Focus and Saccade Tracking (FAST)*, which, through the front-facing camera, approximates when and how the user is looking at the screen, even if users are wearing corrective lenses or glasses.

A high-level description of FAST is as follows. FAST first determines whether or not the user is looking at the screen. If the user is looking, FAST measures the movement of the user's pupil relative to the rest of the eye to track where the user is looking on the screen. In cases when the screen is displaying text (*e.g.*, emails or web articles), FAST detects quick, horizontal jumps. These gaze jumps, or *saccades*, are used as a proxy to determine when the user moves to a new line and to estimate where the user is in a body of text. FAST continually audits its estimate by checking whether the user's observed reading speed is within the expected range (200 – 400 wpm [16]).

Once SwitchBack detects that the user has returned from a distraction by looking back at the screen, it guides the user back to where he or she last left off by highlighting the appropriate region of text. FAST can also be used to enable automatic scrolling when the user reaches the end of the text visible on the screen. This capability can prove very helpful in a number of situations, for example, when a user has gloves on and the capacitive touchscreen will not work.

We evaluated FAST to quantify its performance for attention and saccade tracking, and for supporting reading task resumption. By incorporating information about the user's reading speed, FAST is able to estimate how many lines the user has read in a body of text within a mean absolute percent error of just 3.9%. We evaluated the effectiveness of SwitchBack in helping the user complete their mobile tasks in the presence of external distractions. In a controlled user study, participants, while experiencing distractions, improved their average reading speed during a reading task by 7.7% (roughly 19 words per minute) when using SwitchBack compared to the control condition.

The main contribution of this paper is to demonstrate that SwitchBack can be used in reading applications to facilitate the user resuming the reading of a body of text after attending to an outside distraction. This contribution comes in three parts: (1) the FAST algorithm for detecting when and where the user is looking at the screen using the camera found on commodity mobile phones and without the need for any additional illumination hardware; (2) the ability of FAST to track a user's reading pattern and guide them to the most recently read line of text when they return from another task; and (3) an evaluation of SwitchBack showing that it improves reading speed in situations of divided attention.

RELATED WORK

SwitchBack uses FAST to detect when and how the user is looking at the screen. FAST is related to, and can leverage, any sort of gaze tracking technique. Since SwitchBack is intended to help users with their mobile devices, we also discuss prior research regarding attention tracking for interface interactions.

Gaze-Tracking Technologies and Techniques

The most similar work to SwitchBack is EyePhone [24], a hands-free interfacing system intended for mobile applications that are used while the user is driving a vehicle. EyePhone uses the front-facing camera of a smartphone to monitor the user's gaze on the screen. Rather than tracking relative changes in the user's gaze, however, EyePhone monitors the absolute position of the user's gaze to make selections on the screen. While absolute position provides more information about the user's attention, gaze-tracking accuracy quickly degrades as the phone moves further away from the user's face (<20% accuracy for button selection at ~45 cm). By comparison, our use of relative changes can be easily extracted from the noisy eye-tracking signal, particularly the large ones that occur as the user's eyes travel from one side of the screen to the other. We will demonstrate how relative tracking results in a robust system suitable to use in motion.

Thorough reviews of gaze-tracking have been written by Hansen and Ji [12] and Morimoto and Mimica [26]. We briefly highlight a few innovations and direct readers to their surveys for more detail. Active gaze-tracking systems typically use infrared (IR) light because of the glint that appears when IR light is reflected off of the boundary of the lens and the cornea. Off-the-shelf devices[1,2] and wearable sensors integrated with eyeglasses [35] are normally used to shine IR light onto the user's face. From there, either Pupil Center Corneal Reflection [11] or machine learning [2] is used to learn resulting gaze coordinates. Passive gaze-tracking systems do not rely on any extra hardware, but rather process video and images from the camera using computer vision techniques. Machine learning methods like neural networks [33] have been used to develop a mapping from high dimensional image data in pixel space to low dimensional data in gaze-coordinate space.

Beyond EyePhone, gaze-tracking has not been heavily investigated in mobile computing. Drewes *et al*. [6] devised gaze-based gestures for a mobile device, but utilized an external IR tracker. Commercial entities have perhaps advanced development most. Companies like the Eye Tribe[2] sell active IR-based eye-trackers that are small enough to be mobile device accessories. Samsung gives the appearance of gaze-tracking through their Smart Scroll and Smart Pause features, but these systems actually monitor the orientation of the user's face. Recently, Qualcomm[3] has integrated a passive method of gaze-tracking in the facial processing SDK included on their Snapdragon processor.

[1] http://www.tobii.com/
[2] https://theeyetribe.com/
[3] https://developer.qualcomm.com/mobile-development/add-advanced-features/snapdragon-sdk-android

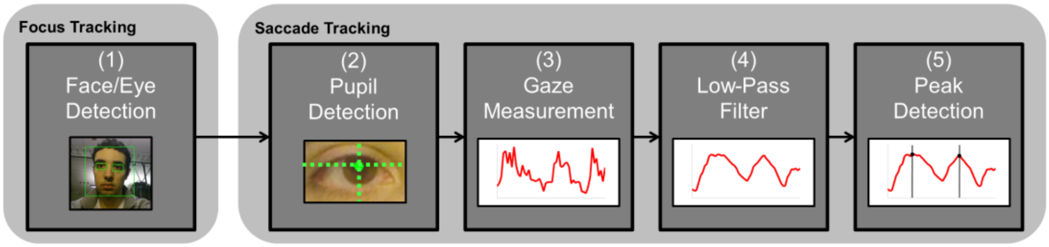

Figure 2. The FAST algorithm. **(1)** The user's face and eyes are detected in order to establish whether the user is focused on the screen or not. If the user is facing the screen, **(2)** the pupils are identified and compared relative to the bounding box of the eye to **(3)** identify the user's gaze in the horizontal and vertical directions. **(4)** These measurements are applied through low-pass filtering and **(5)** peak detection to identify saccades.

FAST uses this Qualcomm library for gaze feature extraction.

Attention and Interface Interaction

The design of SwitchBack's task-resumption feedback draws inspiration from Phosphor [4], a system that modified user interfaces on desktops to highlight changes in widget settings that users may otherwise miss. Similar techniques have been used to draw attention [23] and migrate users between different interface layouts [3]. SwitchBack shows similar transitions when users resume tasks on their mobile devices. Unlike SwitchBack, none of these techniques is aware of users' attention.

Visual fixations have been interpreted as both a measurement of interest and uncertainty [15,16]. Saccades have been used to reveal marked shifts in behavior [9,10]. Scanpaths, or sequences of saccades and fixations, describe how a user searches through an interface [1,10]. Goldberg and Kotval [9] showed that deviation from a normal scanning behavior indicates poor user training or bad layout design. Blink rate [5] and pupil size [22,29] have been used as an indexes of cognitive workload, but may be affected by outside factors like ambient light.

Just and Carpenter [16] monitored the eye fixations of users as they read scientific passages. Their main discovery was that readers make longer pauses when processing loads are greater (*e.g.*, longer words, confusing phrases). They also showed that people read in a saccadic, sequential manner. We use this latter fact to design the reading algorithm underlying our system.

The motivation behind SwitchBack follows closely with that of the work demonstrated in Gazemarks. Kern *et al.* [18] observed that people manage multitasking search situations by using placeholders like fingers or pens. Gazemarks tracks the user's gaze using an off-the-shelf gaze-tracker and provides digital placeholders, in the form of shadowed circles, to guide the user's attention when they switch tasks. We differ from Gazemarks in two ways: (1) We observe that relative gaze changes (*i.e.*, saccades) are better suited for monitoring attention on mobile devices than absolute gaze coordinates. Gazemarks was developed for desktops, which presume a stable environment with

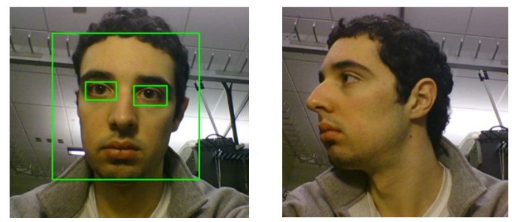

Figure 3. (a) If the user is looking at the screen, his face and eyes are detected. **(b)** When the user begins to look away, his face is either no longer detected or has turned too much for him to be focused on the screen.

large screens and off-the-shelf eye-trackers, thus allowing them to use absolute gaze coordinates to track attention. (2) To motivate our FAST technique, we designed the SwitchBack application. While our interface modification is similar to the one applied in Gazemarks, reading is a use case that Kern *et al*. only briefly mention, but never develop.

THE DESIGN OF *SWITCHBACK*

SwitchBack is built upon our FAST algorithm (Figure 2), which has two major components. The first is *focus tracking*, which determines whether or not the user is facing the screen. The second is *saccade tracking*, which observes the position of the user's pupil relative to the rest of their eye to determine how the user's gaze changes over time. We combine these components within SwitchBack and apply highlighting to aid task resumption.

Focus and Saccade Tracking (FAST)

Focus Tracking

FAST takes frames from the front-facing camera and passes them through the Qualcomm Snapdragon SDK for processing. The SDK detects faces and the orientation of those faces. Both face detection and face orientation are useful for focus tracking. If a face is not detected on the screen, there is no way for the user's eyes to be tracked and the user is probably not attending to the screen (Figure 3b). Even if the user's face is in the view of the camera, the user may not be attending to the screen. In this case, we set bounds on the yaw (*i.e.*, side-to-side angle) of the face according to the typical angle formed between the distance from the user's face to the center of the screen and the

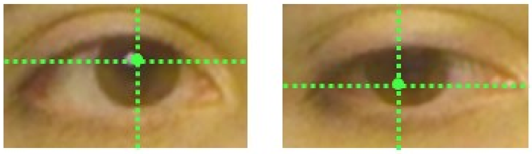

Figure 4. The position of the pupil relative to the rest of the eye changes as the user looks from **(a)** the top left to **(b)** the bottom.

distance from the user's face to the outside edge of the screen. If the user turns his face past these bounds, we infer that the user's attention is directed elsewhere.

Even when the user is facing the screen, face detection may briefly fail due to transient lighting conditions or occlusion by hands. To prevent false triggers indicating that the user has turned away, FAST maintains a sliding window of 2 seconds (overlap of 1/16 fps = 62.5 ms). It only infers that the user's attention has left the screen when the face is not detected for a window's duration. Any distractions shorter than that duration are so brief that the user likely will have not lost their focus from their previous task. As soon as the user is facing the screen again, FAST infers that his attention is back on the screen.

Saccade Tracking

Once SwitchBack detects that the user is looking at the screen, the system tries to detect saccades, or fast eye movements. In cases where the user is reading something on their screen, saccades can be used to estimate the number of lines that the user has read. To detect saccades, SwitchBack tries to detect the user's pupil by finding the darkest portion of his eye.

With the face, eyes, and pupil detected, gaze direction can be quantified using the same geometrical observations that have inspired gaze tracking technology in the past [6,12,26]. Figure 4 shows that the center of the pupil shifts depending on where the user is looking at the screen; we use this fact to form horizontal and vertical features that encode the position of each pupil within their respective eyes. Feature values range from 0 to 1, with 0.5 meaning that the eye is in the exact center along a specific dimension. The features are passed through a first-order infinite impulse response (IIR) filter with a cutoff frequency of roughly 2.5 Hz; this is a low-complexity low-pass filter that suppresses high-frequency noise in time series measurements.

Gaze coordinates on the screen can be inferred either through calibration or knowing the position of the user's head in the camera's field of view. EyePhone [24] divides the bounding box around the user's eye and the screen into a grid and uses a one-to-one grid cell mapping to associate pupil locations to targets on the phone. For our purposes, we do not need to know exactly where the user is looking on the screen. Instead, we are only interested in drastic gaze changes, implying that a saccade has occurred.

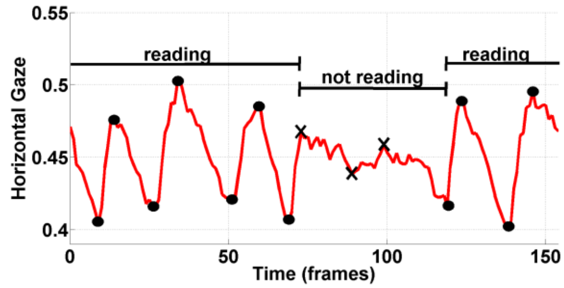

Figure 5. Peaks and troughs are detected throughout the signal, but only indicate saccades when the amplitude of the signal is large. Low magnitude values correspond to when the user is looking to the right, whereas higher magnitude values correspond to looking left.

Figure 5 plots the estimated horizontal location of a user's pupil with respect to time as the user moves their eyes, pauses for a few seconds, and then continues to move their eyes. Low magnitude values correspond to when the user is looking to the right, whereas higher magnitude values correspond to looking left. The signal jumps whenever the user's gaze changes drastically and a saccade occurs. FAST detects these movements in real-time by looking for local optima within a window. Note that saccade detection is agnostic to the amplitude of the signal, so such a system can be prone to noise appearing in flat signals. To prevent this from affecting FAST's line prediction, a detected optimum is only counted as a saccade if the gaze location changes significantly afterwards. The threshold for determining significant changes was determined empirically by iterating through different values and optimizing the results from the data collected in our technology evaluation. In Figure 5, the local optima marked with circles are considered saccades by the algorithm, while the crosses mark some of the local optima that are rejected by the algorithm because the signal exhibits less deviation during those times.

FAST in the Context of Reading

While reading English prose, right-to-left saccades indicate when users begin to read a new line (Figure 5). Between saccades, users' gaze exhibits sawtooth-like behavior as their gaze gradually sweeps across the screen.

Despite the filtering described earlier, there are still cases where noise in the signal and sporadic glances may be inferred as saccades but not necessarily correspond to actual saccades when the user looks at a new line. False positives like these would advance the estimate of the user's most recently read line too far. If face and eye detection fail for a brief period of time, it is also possible that saccades will be missed. False negatives cause the estimate of the user's reading position to lag behind. If there is no mechanism for correction, the errors can accumulate over time and cause frustrating results for the user.

To remedy this issue, we take advantage of the fact that we know the arrangement of the text within the application. In other words, we know how many words[4] appear in each line. Combining this with knowledge about typical human reading speeds (200-400 words per minute [16]), FAST can predict how long it expects a user to read the next line of text, comparing that expected range, <t_{min}, t_{max}> with the measured interval between saccades, $t_{measured}$. There are three possibilities:

1. $t_{min} < t_{measured} < t_{max}$: The measured interval falls within the expected range, so we accept the saccade as a proxy for a new line and increment the line estimate accordingly.

2. $t_{measured} < t_{min}$: The measured interval falls short of the expected range, so we infer that not enough time has passed between line breaks, label that saccade as a false positive, and do nothing to the line estimate.

3. $t_{measured} > t_{max}$: The measured interval was longer than expected, so at least one saccade was likely missed. This means that the user read more text than what appears on the current line. We look ahead in the text, increment the line count, and update t_{min} and t_{max} until the expected range surrounds $t_{measured}$.

If the user looks away, pauses, or backtracks slightly while reading a line, $t_{measured}$ will overestimate the time it takes for the user to read the line. These moments can be segmented out of the time-varying signal, as shown in Figure 5, because they do not share the same sawtooth-like behavior as the reading moments. Therefore, $t_{measured}$ is only calculated using segments when the signal has the sawtooth-like behavior and the user is reading.

There are a couple of further considerations that should be taken into account when incrementing the line count. First, the current line for the user must lie within the screen; *e.g.*, if the first five lines of an article appear on a page, there is no way that the user is reading the sixth line. We use this fact to correct SwitchBack's estimate when the error becomes very large. Another important consideration to make about the text is line length. In Figure 6, for example, a user was able to jump directly to the third line after she read the first. This is caused by the fact that the user was able to read the second line during the right-to-left saccade (marked by the red arrow in Figure 6). We handle cases like this by ignoring lines that span less than a third of the screen (including line breaks) in the aforementioned calculations.

Technology Evaluation of FAST

Prior to conducting our primary user evaluation of SwitchBack, we conducted a technology evaluation to inform our design and quantify the accuracy of FAST. We collected reading data from 8 participants (5 males, 3 females), each of whom volunteered for a 20-minute session. Two of the participants wore glasses during the experiment.

[4] We use "word" to mean a group of five characters, consistent with terminology used in text analysis.

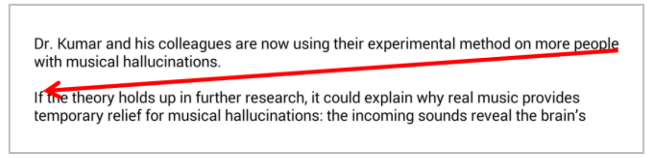

Figure 6. When there is spacing between lines, the user's eyes can follow the arrow and still read the second line.

We collected data using our custom reading application (Figure 1) on a Sony Xperia Z smartphone, which has a 5-inch screen, 1080×1920 pixel display, and 2.3 MP front-facing camera. Participants were asked to hold the smartphone in landscape mode while reading 6 different excerpts of text from the New York Times with size 15 font (roughly 1.8 mm in height) and presented as a single contiguous paragraph. Each article was clipped to 20 lines in length so the user did not have to scroll. We chose to use contiguous text in this experiment to gather sufficient data to estimate the time it takes for the typical user's eyes to travel across the width of the screen. The text was displayed on the screen with black text on a white background. Participants were asked to read 3 articles while standing and 3 articles while walking on a treadmill at a comfortable pace. The study was counterbalanced across conditions. The walking condition was added to evaluate how FAST performs in the presence of extraneous vibrations due to walking. Horizontal and vertical gaze angle were recorded by the application and processed offline using MATLAB.

Line prediction error is defined as the difference between the number of lines in the excerpt and FAST's line prediction. Our results show that FAST achieves a mean absolute prediction error of 0.8 (SD = 1.1) out of the 20 lines of text per trial, translating to 3.9% error. We also found that most of the errors tended to be positive, indicating that FAST overestimated the user's reading position slightly.

To delve deeper into our results, a mixed-effects model analysis of variance was used to analyze the data, with fixed effects for *Walking*, *Glasses*, and *Gender*, and a random effect for *Subject*. (Random effects are appropriate when levels of a factor are not of specific interest, but represent a larger population about which inferences are meant to be drawn [7]. Mixed-effects models are also appropriate for handling repeated measures over the same subjects due to their ability to model covariance in the data [21].)

We expected FAST to be more accurate while participants were standing than walking. Although FAST looks at relative, and not absolute, pupil position, we believed that the shaking of the front-facing camera caused by walking would still affect our performance. However, the average

Figure 7. To support the different conditions, the experimental setup included a treadmill for walking, a lamp for consistent lighting, and a monitor on either side for distractions.

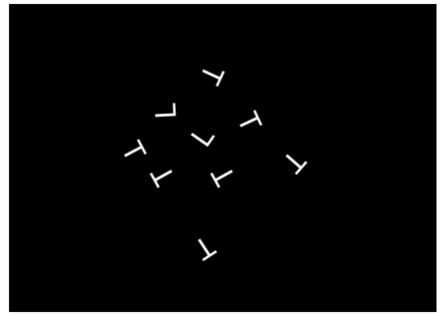

Figure 8. The distraction task required participants to distinguish between rotated T's and L's within a fixed amount of time.

lines-of-error while walking was 0.30 (*SD* = 0.92), and while standing was 0.29 (*SD* = 1.42), so there was no detectable difference ($F_{(1,29.9)} = 0.00$, *n.s.*).

We also expected FAST to perform worse for users who wore glasses because we believed that the computer vision component within FAST might perform poorly. The average error for people wearing glasses was 0.50 (*SD* = 1.83), and the error for those not wearing glasses was 0.21 (*SD* = 0.82), but that difference was also not statistically significant ($F_{(1,4.1)} = 0.52$, *n.s.*).

We did not expect any difference in performance across gender. The average error for males was 0.36 (*SD* = 1.32) and for females was 0.19 (*SD* = 0.98). This difference was not statistically significant ($F_{(1,4.7)} = 0.18$, *n.s.*).

SwitchBack Reading Application

We developed a reading application to guide the user's attention back to where they were last looking in a body of text after attending to an outside distraction (Figure 1). While users are reading, FAST keeps track of where they are looking using saccade tracking corrected with information about the text. FAST detects when users look away from the screen and saves the estimated line a user was last reading. Once a user turns her attention back to the screen, SwitchBack highlights the line that was saved, aiding task resumption.

USER EVALUATION OF SWITCHBACK

We conducted a user study for our SwitchBack reading application with the intent of demonstrating that SwitchBack allows users to more easily resume reading on mobile devices after looking away due to distraction.

Participants

Seventeen participants (9 male, 8 female) ranging from 19 to 52 years old (*M* = 26.7, *SD* = 7.2) were recruited for our study. The participants were evenly distributed between Caucasian, Asian, and South-Asian races. Five participants wore glasses during the study, and all but two of the participants owned and used a smartphone on a daily basis.

Apparatus

Participants used our custom application on a Sony Xperia Z smartphone with a 5-inch capacitive touch screen, 1080×1920 pixel display, and 2.3 MP front-facing camera. The reading application was the same as the one developed for the FAST evaluation study. The only difference was that SwitchBack would be enabled in one condition. When active, SwitchBack guided the user's focus of attention by highlighting a line of text, as shown on the right side of Figure 1. The text was highlighted whenever the user turned away from the screen and then disappeared 5 seconds after they returned to the application. Eight New York Times articles between 500 and 600 words (*M* = 522, *SD* = 18) were chosen for the study; their average readability score was 46.3 according to the Flesch Reading Ease test [19]. A diffused light source subtly illuminated the experiment area, ensuring consistent lighting conditions for all participants; special care was exercised to ensure that the lighting would not affect participants' reading ability.

To study the effect of SwitchBack in the presence of distractions, we devised a secondary distraction task and introduced it as a condition. The task was adapted from an attentionally-demanding task used by Yokoyama et al. [36]. The software for the task was written in C++ and ran on a Windows desktop with two monitors placed on either side of the user (Figure 7). Both screens began completely black. Every 25-40 seconds, one of the two screens played a tone to direct the user's attention away from the phone. The screen then displayed a combination of T's and L's (always 9 in total) with white text and at different orientations (Figure 8). Time intervals, screen selection, letter rotations, and letter selections were all randomized to simulate unexpected distractions and prevent anticipation.

To simulate a natural setting where users might face distractions, participants were asked to walk on a treadmill (Figure 7) for a portion of the experiments. The treadmill was set to a speed of about 1.4 m/s.

Procedure

The procedure was designed to fit in a single 45-minute session. Each session began with a pre-study questionnaire about the participant's mobile device experience and habits. Once completed, participants were introduced to the experimental setup and asked to familiarize themselves with the application. We also informed participants that they would be tested for reading comprehension through a four question multiple choice test at the end of each trial. Participants read a different article per trial to avoid familiarization with the text or encouraging them to skim. The order of the articles was randomized using Latin Squares to ensure that specific articles did not impose a bias on a particular set of conditions.

At the beginning of each trial, the experimenter ensured that the participant began reading with their face within the front-facing camera's field of view when the phone was held in landscape mode. Participants were asked to read each article in its entirety and to scroll whenever they desired. When the distraction task was active, participants were asked to verbally report either the number of T's or L's to the experimenter whenever letters appeared on one of the two outside monitors before returning to the reading; since the number of letters on the screen remained constant, we allowed participants to report the number of either letter. On average, participants experienced 4.4 distractions per trial (*SD* = 1.7) when the distraction task was active.

At the end of each trial, participants were asked to complete a short multiple choice test of reading comprehension and a NASA Task Load Index (TLX) questionnaire [13] to provide feedback concerning their experience.

Design & Analysis

The study was a within-subjects 2×2×2 factorial design. The factors and levels were:

- ***Posture***: *Sitting* and *Walking*.
- ***Distraction***: *Distraction* and *No Distraction*.
- ***Interface***: *SwitchBack* and *Control*.

Posture was the first factor that was counterbalanced. Within each posture, *Distraction* was counterbalanced, followed by *Interface*. Each participant completed every unique combination of conditions, leading to 17×2×2×2 = 136 total trials for the study.

Average reading speed, measured in words per minute (WPM), was the main measure for assessing SwitchBack's performance. To calculate reading speed, we divided the total number of characters in the text read by the time the user spent reading and divide that by 5, the standard for characters per word used in text analysis. Our calculation excludes the time spent completing the distraction task, when applicable.

A mixed-effects model analysis of variance was used to analyze our data, with fixed effects for *Distraction*, *Posture*, and *Interface*, and random effects for *Article* and *Subject* [7,21]. Overall, reading speed was normally distributed according to a nonsignificant Shapiro-Wilk W-test ($W = 0.989$, $p = .383$) [32].

Being ordinal in nature, Likert ratings (1-20) for the NASA TLX instrument were analyzed using the nonparametric Aligned Rank Transform procedure [14,30]. This procedure allows for an analysis of variance (ANOVA) to be used to

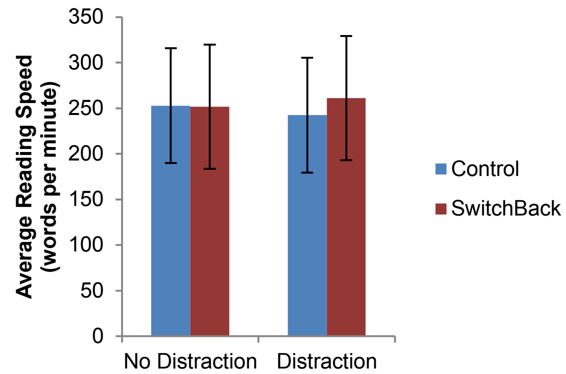

Figure 9. SwitchBack resulted in 19.0 WPM (7.7%) faster reading speeds than the control condition in the presence of distractions. Error bars show standard deviations.

test for main and interaction effects after aligning and ranking the data separately for each effect. Despite using analysis of variance, the procedure is considered nonparametric due to its ranking preprocessing step. The ARTool by Wobbrock *et al.* [34] was used to prepare the data for analysis using the Aligned Rank Transform procedure. Each of the 17 subjects filled out the 6 NASA TLX workload scales (1-20) after each of 8 articles, resulting in 17×6×8 = 816 individual ratings.

RESULTS

Reading Speed

Reading speed results are shown in Figure 9. Across all trials, the average reading speed was 252.0 WPM. This rate is near the lowest quarter of the expected range for typical reading rates, 200 – 400 WPM [16], which we believe can be attributed to the fact that users were asked to read for comprehension, rather than speed.

We looked at the *Interface* × *Distraction* interaction to evaluate the performance of our system. We found this interaction to have only a marginal effect on reading speed ($F_{(1,105)} = 2.93$, $p = .09$). Without our distraction task, SwitchBack never has to modify the interface, so we dug deeper into this trend by looking at the effect of SwitchBack just in the presence of distractions. As we expected, a pairwise comparison shows that SwitchBack significantly increased reading speeds ($F_{(1,107.8)} = 4.61$, $p < .05$) in those cases. This translated to an extra +19.0 WPM increase in users' reading speed when SwitchBack was used in the presence of distractions, a +7.7% increase from the average reading speed across all users. Although the interaction was only a trend, prior work has convincingly argued for the appropriateness of doing *post hoc* comparisons on trend-level interactions [8].

We also gathered findings that confirmed that our experimental treatments did indeed increase workload, as intended. The *Distraction* × *Posture* interaction had a marginal effect on reading speed ($F_{(1,110.6)} = 2.83$, $p = .09$), and in the presence of distractions, walking lowered reading speeds ($F_{(1,111.7)} = 4.01$, $p < .05$).

NASA TLX Questionnaire

It was our hope that through our NASA TLX questionnaires, we would demonstrate that SwitchBack generally decreased perceived workload. While we were not able to detect any significant differences, we found that SwitchBack marginally lessened the mental demand experienced by our users ($F_{(1,112)} = 3.25$, $p = .07$). This finding corroborates the findings concerning improvement in average reading speed; that is, reading speed should increase if mental demand decreases since the user is allowed to focus more on reading.

The NASA TLX questionnaires were more revealing about the effects of our dual focus-of-attention study design. For example, both walking ($F_{(1,112)} = 21.56$, $p < .0001$) and the distraction tasks ($F_{(1,112)} = 21.09$, $p < .0001$) significantly increased frustration for users. Similar trends can be observed across all the other aspects of the survey: mental demand, physical demand, temporal demand, success/failure, and effort.

DISCUSSION

Our goal was to develop a system that eases the user back into a task on their mobile device. To demonstrate this, we developed a SwitchBack reading application that identifies the user's reading pattern through gaze tracking and guides the user back to the proper location in the text after turning away. SwitchBack's FAST algorithm was able to identify the appropriate line to highlight with a mean absolute percent error of 3.9%. In an evaluation of SwitchBack's effect on performance during a reading task with distractions, SwitchBack improved average reading speeds by 7.7% in the presence of distractions.

We focused on smartphones for our user study because smartphones are currently one of the most pervasive devices in the world and we wanted to explore the constraints of working with small screens. SwitchBack is even better suited for devices like the Amazon Kindle, which are made specifically for long-form reading. In fact, we are confident that FAST would have higher accuracy with such devices because larger screens create more smooth and noticeable saccades. By tackling the hardest set of conditions for validating FAST (*i.e.,* a small screen and walking), we informally demonstrated that FAST would work on larger devices; doing the converse would not have been possible.

While evaluating the performance of the FAST algorithm alone, we found that it performed sub-optimally in roughly 15% of the trials. These can be attributed to two causes. The first cause was occlusion when the user unknowingly covered the camera with their thumb. We conducted our experiments in the device's landscape orientation because the Snapdragon SDK did not have the accuracy to support the narrower portrait news article width. Future mobile devices with multiple front-facing cameras, like the Amazon Fire, may alleviate such issues. The second cause relates to situations when participants were unable to tell if their face was within the camera's field of view while reading. We considered adding visual feedback to remedy this issue, similar to the approach taken by Samsung for their Smart Scroll feature [17]; however, we found this to be a distraction in itself that led to extraneous saccades. A wide-angle camera lens would alleviate the issue of the user's face moving out of the camera frame in most cases.

While testing SwitchBack's reading application, we found that some of our users had already developed their own way to keep track of where they were in a large body of text. For instance, one participant told us that she scrolled so that the line she was reading was always at the top of the screen; of course, she could not rely on this once she had scrolled to the bottom of the page. Other participants stated that they kept track of where they were by remembering a key phrase in the article as a "mental bookmark," despite our complicated distraction task. While such bookmarks should impose more cognitive load on the user and impair their ability to complete other tasks, a field study with ecologically valid distractions would be insightful towards examining more realistic cognitive load tradeoffs.

Our use of FAST in the SwitchBack application ignores gaze position in the vertical direction because of its poor accuracy, which may be partially attributed to the short height of the smartphone screen while it is the landscape orientation. In ignoring the vertical gaze position, we were forced to concede a few assumptions involving the advancement of the line count. First, we assume that the user begins reading from the first line of text. We believe this is fair for new bodies of text that the user has not seen previously, but there is a solution for accommodating other starting locations without vertical gaze position. Since SwitchBack is implemented on devices with touchscreens, the user can double tap a word to simultaneously activate the gaze-tracking and notify the system of their starting point. The second assumption we concede in SwitchBack is that the user reads each line in sequential order. This assumption is confirmed by research in psychology that focuses on reading analysis [16], but the possibility still remains that users will go back and reread missed portions of their text. If the saccade is small and occurs quickly after the user reaches a new line, FAST will treat it as a false positive and continue tracking as normal; however, our algorithm does not address jumps when the user goes back *multiple* lines.

Currently, SwitchBack is a one-size-fits-all system (*i.e.*, no training or customization for each user); however, one could imagine a system that learns the device owner's reading habits over time. SwitchBack could begin by checking if the user's reading speed falls within the wide range of typical human speeds, as in our implementation, but then narrow that window to the particular user's reading speed to ensure more reliable line prediction. The performance of SwitchBack could also be improved by taking into account the text being read. Currently, SwitchBack uses word count to estimate the amount of time it expects the user will take to read a line. Lines with larger words and more complicated content require a heavier cognitive load [20], causing the user to spend more time reading them. We examined articles from The New York Times that were of medium cognitive load according to the Flesch Reading Ease test [19]. Accounting for the content of the text on a deeper level would improve SwitchBack's estimation of the user's reading speed and allow our system to scale to text of varying cognitive load (*e.g.,* scientific articles and children's stories) and different reading behaviors (*e.g.*, skimming). Even further, techniques like Kalman filtering could account for the content of the text and personalize predictions.

FUTURE WORK

The SwitchBack reading application modifies a user interface by highlighting where the user was last reading after he or she attends to an outside distraction. There are other reading application modifications that can be applied using FAST. One that was explored, but not tested in our user study, was automatic scrolling once the user reaches the bottom of the text, similar to Samsung's Smart Scroll [17]. Currently, Smart Scroll and Smart Pause only use face detection and orientation to control the screen, but incorporating information about the user's eyes could provide a better experience. Another possible modification, geared primarily towards users with poor eyesight, could be a magnifying glass-like feature that enlarges the current line of text for the user.

Testing the SwitchBack outside of a laboratory setting would validate the application's robustness. We did not evaluate SwitchBack outdoors because of the lack of control over conditions like the number of distractions encountered. We simulated walking with the treadmill to introduce some of the factors that would be met outdoors, but we have yet to test SwitchBack in different lighting conditions. We are confident that applications involving relative gaze changes are more robust to such conditions than applications involving absolute gaze position since saccades may be inferred with missing or incorrect data.

We have used FAST in the context of reading, but we believe that it enables a broader range of applications. For instance, SwitchBack could be used by advertising companies to gauge whether or not their advertisements are engaging to users. FAST can be applied to any combination of text and images so long as the layout of the content on the screen is known.

CONCLUSION

As people become more attached to their mobile devices, the ability to balance attention towards their devices and awareness of their surroundings deteriorates. We have presented SwitchBack, a generalizable system for easing the user's focus-of-attention back into a mobile device task after attending to an outside distraction. To evaluate SwitchBack, we focused on reading applications. We performed a technology evaluation to determine the performance of our Focus and Saccade Tracking (FAST) algorithm and found that we were able to estimate how many lines participants had read in a body of text to within a mean absolute percent error of just 3.9%. We then conducted a user study on our custom SwitchBack reading application. SwitchBack increased participants' reading speeds by 7.7% in the presence of distractions. It is our hope that SwitchBack and FAST will prove useful towards realizing more situationally-aware mobile devices in the future.

ACKNOWLEDGEMENTS

This work was supported in part by the National Science Foundation under grant IIS-1217627. Any opinions, findings, conclusions or recommendations expressed in this work are those of the authors and do not necessarily reflect those of the National Science Foundation.

REFERENCES

1. Aaltonen, A., Hyrskykari, A., and Räihä, K. 101 spots, or how do users read menus? *Proc. CHI '98*, (1998), 132–139.
2. Baluja, S. and Pomerleau, D. *Non-intrusive gaze tracking using artificial neural networks.* 1994.
3. Bartram, L. Can motion increase user interface bandwidth in complex systems? *Proc. Systems, Man, and Cybernetics '97.*, (1997), 1686–1692.
4. Baudisch, P., Tan, D., and Collomb, M. Phosphor: explaining transitions in the user interface using afterglow effects. *Proc. UIST '06*, (2006), 169–178.
5. Brookings, J., Wilson, G., and Swain, C. Psychophysiological responses to changes in workload during simulated air traffic control. *Biological psychology 42*, 3 (1996), 361–377.
6. Drewes, H., Luca, A. De, and Schmidt, A. Eye-gaze interaction for mobile phones. *Proc. Mobility '07*, (2007), 364–371.
7. Frederick, B. Fixed-, Random-, and Mixed-Effects ANOVA Models: A User-Friendly Guide for Increasing the Generalizability of ANOVA Results. *Advances in Social Science Methodology*, (1999), 111–122.
8. Games, P. Multiple comparisons of means. *American Educational Research Journal 2*, 8 (1971), 531–565.
9. Goldberg, J. and Kotval, X. Computer interface evaluation using eye movements: methods and constructs. *International Journal of Industrial Ergonomics 24*, 6 (1999), 631–645.

10. Goldberg, J. and Stimson, M. Eye tracking in web search tasks: design implications. *Proc. Symposium on Eye Tracking Research and Applicaitons '02*, (2002), 51–58.
11. Goldberg, J. and Wichansky, A. Eye tracking in usability evaluation: A practitioner's guide. *The mind's eye: cognitive and applied aspects of eye movement research*, (2003), 537–605.
12. Hansen, D. and Ji, Q. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence 32*, 3 (2010), 478–500.
13. Hart, S. and Staveland, L. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology 52*, (1988), 139–183.
14. Higgins, J. and Tastoush, S. An aligned rank transform test for interaction. *Nonlinear World 1*, 2 (1994), 201–211.
15. Jacob, R. and Karn, K. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind*, (2003).
16. Just, M. and Carpenter, P. A theory of reading: From eye fixations to comprehension. *Psychological review 87*, 4 (1980), 329.
17. Kang, D., Ryoo, S., Kim, J., Kim, C., and Seo, Y. Apparatus and method for detecting speaking person's eyes and face. *US Patent 6,611,613*, 2003. http://www.google.com/patents/US6611613.
18. Kern, D., Marshall, P., and Schmidt, A. Gazemarks: gaze-based visual placeholders to ease attention switching. *Proc. CHI '10*, (2010), 2093–2102.
19. Kincaid, J.P., Fishburne Jr, R.P., Rogers, R.L., and Chissom, B.S. *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel.* 1975.
20. Kintsch, W. and Keenan, J. Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology 5*, 3 (1973), 257–274.
21. Littell, R.C., Henry, P.R., and Ammerman, C.B. Statistical analysis of repeated measures data using SAS procedures. *Journal of Animal Science 4*, 76 (1998), 1216–1231.
22. Marshall, S.P. Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity. *US Patent 6,090,051*, 2000. http://www.google.com/patents/US6090051.
23. McCrickard, D.S., Catrambone, R., and Stasko, J.T. Evaluating animation in the periphery as a mechanism for maintaining awareness. *Proc. IFIP '01*, (2001), 148–156.
24. Miluzzo, E., Wang, T., and Campbell, A.T. EyePhone: activating mobile phones with your eyes. *Proc. SIGCOMM '10*, (2010), 15–20.
25. Monsell, S. Task switching. *Trends in Cognitive Sciences 7*, 3 (2003), 134–140.
26. Morimoto, C.H. and Mimica, M.R. Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding 98*, 1 (2005), 4–24.
27. Nasar, J.L. and Troyer, D. Pedestrian injuries due to mobile phone use in public places. *Accident Analysis & Prevention 57*, (2013), 91–95.
28. Newcomb, A. Texting While Walking Banned in New Jersey Town - ABC News. 2012. http://abcnews.go.com/blogs/headlines/2012/05/texting-while-walking-banned-in-new-jersey-town/.
29. Pomplun, M. and Sunkara, S. Pupil dilation as an indicator of cognitive workload in human-computer interaction. *Proc. International Conference on HCI '03*, (2003).
30. Salter, K. and Fawcett, R. The ART test of interaction: A robust and powerful rank test of interaction in factorial models. *Communications in Statistics: Simulation and Computation 22*, 1 (1993), 137–153.
31. Sears, A., Lin, M., Jacko, J., and Xiao, Y. When computers fade: Pervasive computing and situationally-induced impairments and disabilities. *Proc. HCI International '03*, (2003), 1298–1302.
32. Shapiro, S. and Wilk, M. An analysis of variance test for normality (complete samples). *Biometrika 52*, 3,4 (1965), 591–611.
33. Stiefelhagen, R., Yang, J., and Waibel, A. Tracking eyes and monitoring eye gaze. *Proc. Workshop on Perceptual User Interfaces*, (1997), 98–100.
34. Wobbrock, J.O., Findlater, L., Gergle, D., and Higgins, J.J. The Aligned Rank Transform for nonparametric factorial analyses using only ANOVA procedures. *Proc. CHI '11*, (2011), 143–146.
35. Ye, Z., Li, Y., Fathi, A., et al. Detecting eye contact using wearable eye-tracking glasses. *Proc. UbiComp '12*, (2012), 699–704.
36. Yokoyama, T., Sakai, H., Noguchi, Y., and Kita, S. Perception of direct gaze does not require focus of attention. *Scientific reports 4*, (2014), 3858.