Alex Mariakakis Research Statement

The proliferation of ubiquitous technologies like smartphones and smartwatches is increasing the number of sensors that travel with us at all times. Motion, audio, and visual sensor data can be analyzed to provide suggestions catered to individuals at their convenience. Theses sensors can also operate in the background and intervene when a serious situation arises. **My dissertation work focused on the application of machine learning and computer vision on data from the smartphone's built-in sensors to create mobile apps that improve access to health screening and safety tools.** I work in the areas of ubiquitous computing and human-computer interaction, but my research is highly interdisciplinary; I collaborate with clinicians, designers, and informaticists. Because my work is so diverse, I have the ability to conduct clinical, survey, and prototype-focused user studies.

As my colleagues and I note in an article on the challenges of smartphone-based health sensing [7], uncontrolled environments and device heterogeneity add noise to sensor data, breaking assumptions that can be made in laboratory settings. Reliability is critical when sensors are being used to keep us healthy and safe at all times. Therefore, **my research approach in mobile health (mHealth) often involves making the tradeoff between scalability and reliability through hardware compromises**. My dissertation work revolved around smartphones because of their ubiquity, but I often incorporated inexpensive accessories (e.g., color calibration glasses, 3D-printed box) to account for the various settings where smartphones may be used. My skills and methods could easily generalize to any sensor package, and the findings from my research can be used to inform their designs.

My long-term academic goal is to push the public view of mHealth beyond step counters and heart rate monitors to health-screening technologies. I envision a world where mHealth apps operate in the background of a person's smartphone, continuously updating their understanding of the user's behavior, environment, and physiology. When an app suspects that the user may have a condition, it deploys an explicit task to support or refute that suspicion. For example, if an app begins to passively detect the onset of Parkinsonian tremor, it would deploy a battery of motor coordination tasks to investigate further. **My vision for future health technologies applies the Bayesian diagnostic process that is often used in clinics but is currently missing from mHealth.** Below, I describe how my research has represented the first step towards this vision, producing explicit tests of health and safety.

PH.D. RESEARCH

Smartphone-based Health Sensing [3,4,8]

Currently, healthcare is most effective when providers see patients in their clinics. Ubiquitous smartphone-based health sensing can help patients be proactive in their own care, lessen clinical resource burdens, encourage adherence to treatment regiments outside of clinics, and provide relevant interventions at the most appropriate time. My broad skillset in computer vision and machine learning can apply to many health-related domains, but my work in this space focuses on symptoms that manifest in the eyes.

One attribute I have explored is the pupillary light reflex (PLR)—the manner in which a person's pupils react to a light stimulus. Emerging evidence has shown that an impaired PLR may be a useful biomarker of traumatic brain injuries (TBIs) like sports-related concussions. Professional sports teams can afford to have trained physicians on the sidelines deciding whether or not an athlete who has suffered a concussion can return to their game; however, teams with less funding must rely on a school nurse or volunteer who does not have the same expertise or tools at their disposal.

PupilScreen [4] is a smartphone app that achieves similar functionality as a clinical pupillometer—a device that uses an infrared camera to measure pupil size over time—at a fraction of the cost. PupilScreen constricts a person's pupils with the smartphone's flash and records the pupils' response through the camera. The video is processed using a fully convolutional neural network, producing per-pixel pupil segmentation for each frame. Those measurements are collated with timestamps to generate the same pupil diameter-over-time curves that a pupillometer would (Figure 1, bottom). Keeping the lighting stimulus consistent across scenarios is important for ensuring that the test is repeatable. Compromising scalability for the sake of consistency, I designed a 3D-printable box that blocks out ambient light and controls the distance between the smartphone and the person's eyes (Figure 1, top). We evaluated PupilScreen through a study on 42 individuals with a normal PLR to ensure that the algorithm could properly capture non-trivial pupil dynamics. PupilScreen was able to track pupil diameter with a median error of 0.30 mm, which is between the accuracy of a pupillometer (0.23 mm) and human observers (0.50 mm). We tested PupilScreen on six patients with TBI in a short follow-up study. When

Figure 1. (top) The PupilScreen box controls lighting while the pupillary light reflex is being measured. **(bottom)** An example of pupil constriction.

clinicians were given PupilScreen's output alone, they were almost perfect when distinguishing responsive and unresponsive pupillary light reflexes.

Another symptom I have explored during my dissertation work is jaundice—yellow discoloration of the skin and eyes caused by a buildup of bilirubin in the bloodstream. Adults can become jaundiced as a result of complications of the liver or pancreas, such as alcoholism and pancreatic cancer. Bilirubin is only measured through a blood draw when a clinician suspects a relevant disease. The manifestation of jaundice can lead to a blood draw, but it is only obvious in severe states. Accessible jaundice detection can help undiagnosed individuals catch serious conditions early and enable diagnosed individuals monitor their own condition.

BiliScreen [3] analyzes the color of a person's sclera through a smartphone photograph. The algorithm starts with sclera segmentation, which is non-trivial since jaundice changes the sclera's color profile. BiliScreen segments the sclera using an automated variant of GrabCut [9]. The sclera's color is transformed into a feature vector that is fed into a random forest regression model that estimates a person's bilirubin level. Repeatable color measurement is of the utmost importance in BiliScreen, yet pictures can be taken in various lighting conditions. Here again, I developed two potential accessories that trade scalability for reliability. The first option was the same head-mounted box that was used for PupilScreen. The second option was a pair of paper glasses with colored squares around their rims that serve as known color references for calibration (Figure 2). We found that BiliScreen performed better with the box than with the glasses in a 70-person clinical trial, achieving a sensitivity of 89.7% and a specificity of 96.8%.

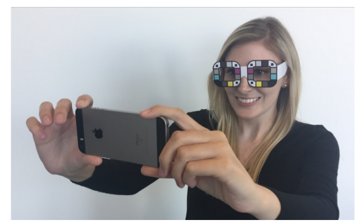

Figure 2. The BiliScreen glasses include squares along the rims for color calibration.

Public Safety [5,6]
I examine issues of public safety from the perspective of situational impairments [10]—contextual phenomena that impede a person's ability to perform a task. Smartphones are often involved in situational impairments because we carry them with us at all times, but smartphones can also detect these impairments through their built-in sensors.

One situational impairment I investigated is inebriation. Portable breathalyzers are the status quo for measuring inebriation, but they are typically used after a drunk driver has been caught—rarely to prevent a drunk person from driving in the first place. Ubiquitous inebriation testing connected to a smart-locking mechanism could help prevent impaired people from getting behind the wheel. **Drunk user interfaces** [6] (DUIs) measure a combination of task-specific performance metrics and sensor-derived features to estimate a person's blood alcohol level (BAL). For instance, performance on a typing task can be characterized by typing error rate and how off-center the user hits each intended key. Features within and across different DUIs are processed by a random forest regression to estimate BAL. It is important to note that DUIs assess side effects of inebriation rather than BAL directly; other factors like fatigue and learning can also affect a person's performance on DUIs. Because of these confounds, we collected data from 14 participants in a 5-day longitudinal study. Participants used DUIs at different BALs, but the same time of day to control for fatigue and account for learning. We were able to estimate BAL with a mean absolute error of $0.005\% \pm 0.007\%$. As a point of comparison, the breathalyzer used as a ground truth had a self-reported tolerance of 0.005%.

Although smartphones are a great mechanism for addressing safety, they can sometimes be the cause of safety issues. For example, a pedestrian who is reading email on their smartphone must balance their focus between their smartphone and the physical space in front of them, imposing cognitive load that can lead to frustration or a pedestrian accident. **SwitchBack** [5] alleviates cognitive burden by allowing the user to resume their smartphone task more efficiently. SwitchBack uses the front-facing camera to track the user's attention and gaze. Although precise gaze-tracking is difficult in mobile contexts with small screens, larger gaze direction changes (saccades) are more robust to detect. With that in mind, we targeted a common activity where saccades are prominent: reading. As a person reads from line-to-line, their gaze occasionally jumps from right to left. SwitchBack counts how many lines the user has read, using information like typical reading speeds, rough vertical gaze position, and the text's layout to adjust that estimate. When the user looks away and then returns their gaze back to the screen, SwitchBack highlights the last line of text the user read

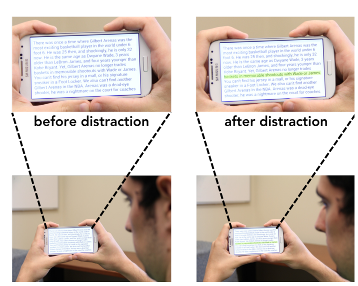

Figure 3. SwitchBack guides a person's attention back to an article by counting the number of lines read.

to guide their attention back to where they left off (Figure 3). To simulate a common pedestrian scenario, we conducted a 17-person study where participants were asked to read news articles and perform a distraction task while walking on a treadmill. We found that SwitchBack improved average reading speed by 7.7% in the presence of distractions, demonstrating that participants were able to attend their surrounding environment and make quicker transitions between contexts without significant cognitive load.

FUTURE RESEARCH PLANS

Achieving my vision for the future of mHealth requires answering a number of questions: **Who** should use a health-screening tool and **when** should they use it? **How** does a person properly use a health-screening tool without supervision? **Why** would a person use a health-screening tool in the first place? I outline how I plan on addressing questions like these below.

Projecting the Acceptance of mHealth Technologies

mHealth technologies are only useful to individuals who are willing to use them and are able to react to the feedback in a logical manner. Testing hypotheses around adoption and acceptability is hard to do when a technology is not mature enough to be deployed. Developers and designers are encouraged to employ rapid prototyping methods to project the reception of their technology ahead of time, but existing methods are not designed to handle some of the unique aspects of mHealth technology. For instance, many health-screening tools have a decision threshold that dictates the tool's sensitivity and specificity. Algorithm developers often set the threshold to maximize overall accuracy, but false positives and false negatives have very different repercussions in the context of diagnostics. Understanding questions like these requires perspectives from computer scientists, human-centered designers, and psychologists. One way I plan to support the investigation of such questions is through a survey instrument and accompanying analysis built around the Health Belief Model [2], a psychological model for health behavior change. The survey instrument would probe people's mental models around health behavior change through hypothetical scenarios, and structural equation modeling would be used to untangle the causal relationships between concepts. With this method, researchers could conduct experiments and extract usable evidence that would inform the design of future mHealth technologies.

Moving from In-Lab to At-Home Testing

When developing a new health sensing technique, researchers often have control over how the data is collected. Unfortunately, first-time users may not be aware of such factors or know when their data satisfies all of the assumptions that underlie a given algorithm. One way to address this issue is to implement real-time algorithms that guide users through the data collection process. For example, I have been advising a junior graduate student who is creating a smartphone app that leverages image processing to guide users to take high-quality photographs of rapid diagnostic tests. Beyond procedural controls, I am interested in identifying machine learning techniques that can be used to create models that generalize across contexts. For example, discriminative adversarial networks can be trained to reward information about a target measure while penalizing information about the data collection environment. Such an approach would not only help algorithms work across contexts but could also potentially obviate the need for device-specific calibration.

Combining Explicit and Ambient Sensing

My long-term research goal involves exploring how continuous ambient sensing can be used to identify the optimal moments for momentary assessments or interventions. Continuous sensing from smartphones (e.g., social network, motion, GPS) or the environment (e.g., temperature, local disease prevalence) can be used to update a Bayesian prior that estimates the likelihood that a person has a given condition. When that prior is sufficiently high, an explicit test can be deployed to support or refute the system's decision, thereby creating a posterior probability. As a postdoctoral researcher, I have already begun fostering three collaborations for exploring such a paradigm: (1) The National Institutes of Health (NIH) periodically runs studies where patients are purposefully infected with influenza in order to study how our bodies respond over time. I am leading a collaboration between Dr. Shwetak Patel's lab and NIAID so that we can study potential ambient indicators of influenza in a controlled environment. (2) While I studied how explicit testing could be used to identify the side effects of inebriation in my DUI work, Dr. Anind Dey studied passive smartphone-based indicators of inebriation [1]. Together with clinicians from the Department of Psychiatry, I led a grant submission to the National Science Foundation's Smart and Connected Health program on a holistic system that would combine our approaches to identify incidences of heavy drinking. (3) At Sage Bionetworks, I have been working on an algorithm that operates on the background of a person's smartphone to identify opportunities when their gait quality can be assessed. This will eventually be incorporated into the NIH's All of Us Research Program, a historic effort to gather data from one million people living in the United States to accelerate research and improve health.

1. Sangwon Bae, Denzil Ferreira, Brian Suffoletto, Juan C Puyana, Ryan Kurtz, Tammy Chung, and Anind K Dey. 2017. Detecting Drinking Episodes in Young Adults Using Smartphone-based Sensors. *Proc. IMWUT '17* 1, 2: 1–36. https://doi.org/10.1145/3090051
2. Nancy K Janz and Marshall H Becker. 1984. The Health Belief Model: A Decade Later. *Health Education Quarterly* 11, 1: 1–47. https://doi.org/10.1177/109019818401100101
3. Alex Mariakakis, Megan A Banks, Lauren Phillipi, Lei Yu, James Taylor, and Shwetak N Patel. 2017. BiliScreen: Smartphone-Based Scleral Jaundice Monitoring for Liver and Pancreatic Disorders. *Proc. IMWUT '17* 1, 2: 20:1--20:26. https://doi.org/10.1145/3090085
4. Alex Mariakakis, Jacob Baudin, Eric Whitmire, Vardhman Mehta, Megan A Banks, Anthony Law, Lynn McGrath, and Shwetak N Patel. 2017. PupilScreen: Using Smartphones to Assess Traumatic Brain Injury. In *Proc. IMWUT '17*, 81:1--81:27. https://doi.org/10.1145/3131896
5. Alex Mariakakis, Mayank Goel, Md Tanvir Islam Aumi, Shwetak N Patel, and Jacob O Wobbrock. 2015. SwitchBack: Using Focus and Saccade Tracking to Guide Users' Attention for Mobile Task Resumption. In *Proc. CHI '15*, 2953–2962. https://doi.org/10.1145/2702123.2702539
6. Alex Mariakakis, Sayna Parsi, Shwetak N Patel, and Jacob O Wobbrock. 2018. Drunk User Interfaces: Determining Blood Alcohol Level through Everyday Smartphone Tasks. In *Proc. CHI '18*, 1–13. https://doi.org/10.1145/3173574.3173808
7. Alex Mariakakis, Edward Wang, Mayank Goel, and Shwetak Patel. 2019. Challenges in Realizing Smartphone-based Health Sensing. *To appear in IEEE Pervasive Computing*, 1–10.
8. Alex Mariakakis, Edward Wang, Shwetak N Patel, and Joanne C Wen. 2016. A Smartphone-based System for Assessing Intraocular Pressure. In *Proc. EMBC '16*.
9. Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. 2004. Grabcut: Interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics (TOG) '04* 23, 3: 309–314. Retrieved May 21, 2016 from http://dl.acm.org/citation.cfm?id=1015720
10. Andrew Sears, Min Lin, Julie A Jacko, and Yan Xiao. 2003. When computers fade: Pervasive computing and situationally-induced impairments and disabilities. In *Proc. HCI International '03*, 1298–1302.