Captivates: A Smart Eyeglass Platform for Across-Context Physiological Measurement

PATRICK CHWALEK, Responsive Environments, MIT Media Lab
DAVID RAMSAY, Responsive Environments, MIT Media Laby
JOSEPH A. PARADISO, Responsive Environments, MIT Media Lab

Fig. 1. Captivates System

We present Captivates, an open-source smartglasses system designed for long-term, in-the-wild psychophysiological monitoring at scale. Captivates integrate many underutilized physiological sensors in a streamlined package, including temple and nose temperature measurement, blink detection, head motion tracking, activity classification, 3D localization, and head pose estimation. Captivates were designed with an emphasis on: (1) manufacturing and scalability, so we can easily support large scale user studies for ourselves and offer the platform as a generalized tool for ambulatory psychophysiology research; (2)

Authors' addresses: Patrick Chwalek, chwalek@mit.edu, Responsive Environments, MIT Media Lab; David Ramsay, Responsive Environments, MIT Media Laby; Joseph A. Paradiso, Responsive Environments, MIT Media Lab.

2474-9567/2021/9-ART93
https://doi.org/10.1145/3478079

robustness and battery life, so long-term studies result in trustworthy data individual's entire day in natural environments without supervision or recharge; and (3) aesthetics and comfort, so people can wear them in their normal daily contexts without self-consciousness or changes in behavior.

Captivates are intended to enable large scale data collection without altering user behavior. We validate that our sensors capture useful data robustly for a small set of beta testers. We also show that our additional effort on aesthetics was imperative to meet our goals; namely, earlier versions of our prototype make people uncomfortable to interact naturally in public, and our additional design and miniaturization effort has made a significant impact in preserving natural behavior.

There is tremendous promise in translating psychophysiological laboratory techniques into real-world insight. Captivates serve as an open-source bridge to this end. Paired with an accurate underlying model, Captivates will be able to quantify the long-term psychological impact of our design decisions and provide real-time feedback for technologists interested in actuating a cognitively adaptive, user-aligned future.

CCS Concepts: • **Hardware → Sensor applications and deployments**; **Sensor devices and platforms**; PCB design and layout; Signal integrity and noise analysis.

Additional Key Words and Phrases: smart eyeglass, robust ambulatory sensing, face temperature, blink sensing, localization

ACM Reference Format:
Patrick Chwalek, David Ramsay, and Joseph A. Paradiso. 2021. Captivates: A Smart Eyeglass Platform for Across-Context Physiological Measurement. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3, Article 93 (September 2021), 32 pages. https://doi.org/10.1145/3478079

1 INTRODUCTION

If we can measure the experiential quality of our work, our entertainment, our education– the mental experiences of deep engagement, stress, alertness, and emotion– we can dissect the influence of our technology and infrastructure on our lives. This analysis would catalyze our ability to actuate our tools and our environments in support of deep, effortless engagement in the tasks we wish to prioritize and otherwise react to our internal state.

Unfortunately, measuring the mental aspect of experience is difficult. Historically social psychologists have studied this using questionnaires, and this remains the only true window into another's inner world. However, psychophysiological modeling has emerged as a new possibility for continuous and unobtrusive approximation of a user's mental experience. Physiological signals have shown strong correlations to mental experience in isolated contexts, like concerts and classrooms, where group data can be easily cross-referenced against emotional stimuli. We believe there is a powerful opportunity to take the insights that have been generated in specific settings and generalize them across real-world contexts. We've identified several key challenges with that translation that we seek to address and overcome with this project.

In this paper, we present Captivates, a smartglasses system designed for long-term, in-the-wild psychophysiological monitoring at scale. Captivates integrates many underutilized physiological sensors in a streamlined package aimed at estimating attention and cognitive state. It was designed with an emphasis on (1) manufacturing and scalability, so we can support large scale user studies for ourselves and outside collaborators without much administrative overhead, (2) robustness and battery life, so we can perform long-term measurement and capture an individual's entire day without interruption, and (3) aesthetics and comfort, so people can wear them across their naturalistic daily contexts without notice, avoiding self-conscious changes in attitude and behavior. These tenants make the Captivates design process closer to a full product design process than a standard research prototype.

We believe this model for scalable, product-like design will become increasingly important in the near future for the fields of human-computer interaction and applied psychology. In this instance– with our focus on psychophysiological monitoring– we believe the Captivates project will serve as a powerful tool for researchers to investigate the links between real-world physiology and phenomenological experience, for designers interested

in grounding the decisions with a model of long-term psychological impact, and for technologists looking to automate supportive systems that must react in real-time to the mental experience of their users.

2 MOTIVATION

There are several studies that suggest technology is powerfully affecting our attention, our focus, and our relationships in ways that are largely negative but difficult to measure. Killingsworth and Gilbert showed the quality of our engagement in the activities we do, rather than the activities themselves, is a better predictor of happiness [44]. This engagement in our learning, our work, our entertainment, and our craft is central to human experience. Designing technology that structurally improves our lives starts with our understanding of these intangible aspects of our mental lives.

In the world of psychology, the standard toolbox for studying these phenomena includes surveys, diaries, and questionnaires. For all but the most ardent behaviorist, these techniques are the only way to collect 'ground truth' information about another's mental experience. Unfortunately, these approaches have a few problems: (1) they require retrospective self-assessment, which we know is biased in complex ways (a simple heuristic is the 'peak-end' rule, in which memory favors salient moments and recent events), and (2) they require interruption, which undermines the experiential quality of engagement we are interested in measuring. Because of this, psychophysiological modeling has become an important tool to approximate internal experience. With this approach, we're able to back out continuous estimates of mental experience from physiological indicators.

Psychophysiological modeling has been used in many domains quite successfully. Skin conductance measurements have been used to measure emotional engagement during live performance [48, 74], alertness in the classroom [29], and stress in call centers [38]. EEG and nose temperature can detect stress or alertness in high-pressure situations like test-taking or air traffic control [11, 17]. There are many studies that show correlations between stimuli of various types– fear, shock, sadness, happiness, cognitive load, attentional state, alertness, stress, and anger– with physiological measurement.

These results are very encouraging, but many issues remain with psychophysiological modeling in current practice if we want to generalize these findings usefully. Some of the main issues include:

- **Laboratory settings don't translate to the real world.** Kurt Lewin, one of the founders of Social Psychology, is known for promoting 'action research', and most discussions of the limitations of modern experimental practice start with his theories. Though these limitations are widely acknowledged, the vast majority of social psychology research is still based on controlled laboratory experiments [60]. Laboratory studies frequently use instrument-grade equipment with supervision; they typically have individuals sit still, acclimate to the temperature, and rest for a few moments before a trial. Participants are typically not moving, talking, eating, or breathing realistically. Once data is collected, in many cases it is pruned by hand with knowledge of where responses to the stimuli in question should appear. Even with many of these practices in place, it can still require averaging over many participants to have confidence in the physiological measure. Because physiology is so heavily influenced by movement, diet, breathing, temperature, and activity level, there are many additional noise sources to consider in real world applications. Instrument grade equipment and standard data processing pipelines also don't directly apply in naturalistic settings.
- **Standard techniques are limited in capturing important dynamics of user experience.** Laboratory experiments often attempt to isolate the part of the physiological signal that is due to a specific stimulus. We test one stimuli in one environment on many users, randomizing out all aspects of their experience that are individual to them. While this is a valid approach, it doesn't provide a sense of the relative importance of the individual and their psychological inertia compared with external stimuli. If a person comes into the lab after having a bad day, feeling stressed about performing well in the study, and then sees a happy

picture, their internal experience is likely dominated by their current mental state and their psychological inertia as opposed to the isolated stimuli.
If we treat someone's mental experiences and behaviors as a function of their state (S) and their environment (E), standard experimental psychology isolates the impact of some attribute of E, which in many cases may be marginal. To learn about S and the dynamics between S and E we need to study across environments. Our ability to create evidence-based, powerful and engrossing tools/experiences (E) that minimize the psychological inertia of the individual– as well as our ability to support self-disciplined mental states (S) that are impervious to the outside context– depend on a systematic analysis of the dynamics of these two things together in natural contexts.

- **HCI research incentives don't support large-scale psychophysiological datasets.** There are incentives in Academia that run against open, large-scale data collection. Creating a prototype is time-consuming, and the result is frequently fragile and difficult to use. Collecting even a small amount of data can be a significant undertaking. Sharing that data has limited utility– it is typically a small one-off dataset that serves a specific hypothesis, there are few additional practical insights to be had, and the prototype is not easily extended to new contexts. Furthermore, data cleaning is time consuming, provenance is frustrating, and there is no incentive to invest extra work simply to have your data scrutinized.
  In machine learning, Stanford professor Fei-Fei Li invested heavily on high quality data collection with ImageNet [28], and opened it to the broader community. This resource enabled large scale DNN training, leading to AlexNet and sparking the deep learning revolution [13]. Mental states are far more stochastic, contextual, and difficult to model than images of objects. Reasonable estimate of mental experience across context will require an even larger set of fundamentally trustworthy data, and a similarly large community focused on modeling it. There must be a similar effort to separate out data collection and mental state modeling to underpin HCI and Social Psychology research.

Captivates is a tool that addresses the major issues listed above. It is a platform to study and translate the principles we see in laboratory studies into the real world, focusing on an individual across natural contexts. It minimizes its presence to preserve the normal course of life as much as possible. It serves as an open platform for psychophysiological research that can be used to collect large amounts of data, and is robust and scalable and easy for others to deploy in their work.

There are many first-order and second-order challenges we hope to address with this project– the first is simply tackling data size and quality in a fundamentally different way. We are collecting realistic data across a user's day, and incentivizing others to also help scale data collection. We're also focusing on diverse, understudied physiological markers like face temperature, head pose, head motion, and blinking. These signals have shown promise in laboratory settings as cognitive indicators– they are partially consciously controlled, and thus provide complementary insight into attention and behavior– but they are poorly understood, especially in naturalistic contexts. Additionally, head-worn IMUs can work effectively as a ballistic heart and respiration rate sensor, which should enable an approximation of these physiological signals with the Captivates platform as well [37]. Captivates is sufficient as a standalone system to study these signals, but can also be combined with a larger fusion of wearable sensors, for instance skin conductance or heart rate measured at the wrist, chest, or finger. We anticipate this will lead to more robust psychophysiological modeling.

If we succeed in building strong, generalized psychophysiological models, Captivates becomes more than a research tool– it will provide feedback for designers interested in understanding how their choices effect the mental experience of their clients; it will provide population-level psychological insight for sociologists; and it will serve as an integral part of responsive systems that need to understand and react to the mental state of their users. The democratization of this kind of psychophysiological modeling would mark a paradigm shift in both psychology and design.

3 RELATED WORK

3.1 Non-contact Physiological Measurements related to Cognitive State

We seek to explore the space of sensor systems that do not require direct contact to the skin, a physically less-invasive technique that results in a more natural user experience and a more robust system for measurements on-the-go. Contact-based sensor techniques are a popular method for measuring physiological signals, but these techniques present challenges in dynamic environments. Contact-based methods include electrocardiogram (ECG) for measuring the heart, electroencephalogram (EEG) for measuring the brain, electrooculography (EOG) for measuring eye movement, and electrodermal activity (EDA) sensing for measuring changes in skin conductance. All of these methods have been applied to cognitive state estimation [33, 45, 47, 68, 77] typically under controlled conditions. Oftentimes, researchers struggle with noise artifacts from internal and external sources [25, 49, 73]. Contact sensing is not ideal for measurements throughout a person's dynamic day because of the mechanical skin-electrode coupling, which is highly susceptible to noise artifacts from user motion [21]. By focusing on non-contact techniques, we have greater flexibility for fit and comfort, and our research can be generalized to systems that are not wearable and do not require physical skin contact.

3.1.1 Face Temperature. As summarized in Table 1, several studies have found that our face temperatures vary with internal state changes (e.g., fear, joy, anxiety, etc.). For example, in [11], it was observed that nose temperatures decrease with the increase of cognitive load due to blood flow restriction when autonomic nerve activity increases. Cognitive load was modulated using a series of reading comprehension and Stroop Tasks (i.e., matching color of a word with the word itself) at varying difficulty. Thermal camera measurements in that study required a fixed distance and perspective of the user; for mobile use, similar resolution would require multiple thermal cameras in each location, is prone to occlusion, and is cost prohibitive.

Table 1. Overview of the Direction of Temperature Variation in the Considered Regions of Interest Across Emotions from [40]

Emotions / Regions	Stress	Fear	Startle	Sexual arousal	Anxiety	Joy	Pain	Guilt
Nose	↓	↓		↑		↓		↓
Cheeks			↓					
Periorbital			↑	↑	↑			
Supraorbital			↑		↑			
Forehead	↓↑	↓		↑	↑		↓	
Maxillary	↓	↓	↓				↓	↓
Neck–carotid			↑					
Nose	↓							
Tail		↓					↓	
Fingers/palm		↓					↓	
Lips/mouth				↑				

An alternative is to fix a face temperature sensing device onto the user. In [51], the researchers fixed a passive infrared radiation sensor (i.e., thermopile) to a tethered, glasses-like wearable to measure nose and forehead temperatures to control for movement. Forehead temperatures are used as a baseline since it is believed to remain stable relative to internal temperature, and face temperature varies with other internal and external events (e.g., convection due to wind). However, forehead temperature was demonstrated to increase with increased cognitive load in [11]. [55] suggests the temple region provides a better estimate of a person's internal body temperature given its proximity to the large superficial temporal artery. Thermometry based on this assumption is used in hospitals, as it approximates the accuracy of more invasive rectal techniques– the most accurate method for measuring internal temperature in a clinical setting [35].

3.1.2 Blink Rate and Eye Gaze. As described in [66], there are four types of eye blinks: reflex blinks, voluntary blinks, non-blink closures, and endogenous blinks. Endogenous blinks– blinks that are not consciously initiated or due to any "identifiable eliciting stimulus"– are theorized to relate to information processing, cognitive load, and task demand.

In [49], test subjects reduced their blink rate under higher cognitive load. Furthermore in [52], eye blink rate correlated with task difficulty over arithmetic tasks (blink rate decreases as the difficulty of the task increases).

Others have studied blink rate as they relate to creativity or emotional shifts, and blink magnitude has been linked to attentional control and alertness [53]. Moreover, spontaneous blink rate drops precipitously when using screens [16] and blink patterns change in response to visual task demands [39]. There has also been significant effort to relate blink rate with dopamine dis-regulation and the study of addiction, though evidence for this specific theory appears weak [26].

The evidence suggests that blink rate and blink magnitude are both powerfully tied to mental experience and poorly characterized. While it's uncommon to find blink rate sensors in existing wearables, a few have been reported in the literature. In [27], the authors detect blink rate and magnitude based on changes in near infrared reflectance using a near infrared diode/phototransistor pair with 85% accuracy.

3.1.3 Head Motion Dynamics. Head motion is an indicator of cognitive effort and attentional control. Cognitively strenuous tasks result in less head movement, and patterns in the data could be used to discriminate common tasks and moments of task shift [50]. Subtle head motion dynamics can predict ADHD in infants [75]; it has been used to measure attention in classrooms and in meetings [57, 72]. Head motion has also been used to predict speech prosody and heart rate [18, 24]; mirroring of head motion between patient and psychotherapist can also predict patient outcomes [58].

Gaze is also a powerful indicator in attention research. Apart from directly linking head dynamics to cognitive state, Stiefelhagen et al. were able to estimate gaze direction from head pose with 89% accuracy [67]. Given some a priori knowledge of the environment, it is possible to predict fixation points using head pose and 3D location; traditionally, this is accomplished with computationally demanding systems like camera-based eye tracking, which can also result in very awkward wearable systems.

3.2 Collecting Physiological Data for Cognitive Analysis in Non-Laboratory Settings

Contact-based measuring techniques for cognitive state estimation are popular, especially for the consumer market. Several products exist that claim to help improve your attention or learning, and aid in relaxation in natural settings. Muse [4] is an EEG device that is designed to help train in meditation, but also claims to "make it easy to access and use brainwave data, inside and outside the laboratory and in real world environments." Muse has made practical trade-offs for real-world use than usual EEG systems, with dry electrodes that do not require additional conductive gel for better coupling [32], but motion artifacts remain. Regardless, Muse no longer directly supports access to the raw data [63]. There are other open-source methods of collecting raw EEG signals, but most of these designs aren't productized for on-the-go applications [5].

There are several commercial smart eyewear devices that include access to sensors data. Google Glass is one such design that has gained traction across the academic spectrum, from analyzing head motion and blink frequency for activity recognition [42], augmented-reality based indoor navigation [59], and even surgical applications [76]. Google Glass was successful as a research platform because it was the first scalable platform that could survive a user's regular day, and it provided a software development kit (SDK). Unfortunately, consumer adoption remains elusive and its design is not socially acceptable. Recent consumer smartglasses attempt to fix this shortcoming, taking their design in a glasses-first direction. Vue [9] is a pair of glasses for activity tracking that offer wireless bone conduction audio for discrete listening. They are designed to look like your average pair of glasses, with all the electronics densely embedded within the plastics. Focals by North [2] takes a similar

approach by offering activity tracking and a projected display onto one of the lenses for notifications and short information intake. Unlike Google Glass, the display in Focals is not visible from the outside world, allowing the glasses to retain a traditional design aesthetic. Both the Focals and Vue are sensor-rich systems and follow a more consumer-friendly design approach but, unfortunately, offer no ability for researchers to grab any of the raw sensor signals or to interface their own sensors.

To the best of our knowledge, J!NS MEME [10] is the only commercial platform that takes a design-first approach for smart eyewear while also benefiting researchers with an SDK (though their software applications are not currently supported outside of Japan). The platform includes a 3-axis gyroscope and accelerometer (420 USD) and optional EOG sensor (1050 USD). The EOG-version of the device has been used by researchers for monitoring fatigue and drowsiness levels, but has been criticized for poor sensor performance and a lack of noise robustness (i.e. when talking or moving) [62, 69]. As explained in Section 3.1 and further supported by these studies, contact-based sensing includes additional data challenges for ambulatory measurements. Furthermore, available gyroscope and EOG sensor data are processed through proprietary algorithms before exposed to the researcher. J!NS MEME is a clear step in the right direction, but its lack of support outside of Japan, proprietary obfuscation of raw data, and high price make it difficult for researchers to adapt it to their custom applications.

More general open-source hardware acquisition platforms do exist in literature that allow researchers to interface multiple physiological sensors. In [20], a device is presented that uses a non-contact radiation measurement for temperature and electrode-based techniques for ECG and EDA. Their results are promising, however their system was designed for lab use (non-wearable and tethered), similar to OpenBCI. This is often the case for these platforms— if they aren't commercialized, there is little incentive to complete a robust design for other researchers to use.

4 DESIGN PRINCIPLES

Given our motivation to build novel physiological sensing into a platform that would enable large-scale, long-term study of natural contexts, we distilled our objectives down into a set of design principles. These considerations were weighed with the help of various traditional and smart eyeglass manufacturers and product designers in the United States, Shenzhen, China, and Seoul, South Korea.

4.1 Scalability

Given the significant time investment this platform demanded, we knew it was important to create a system that could serve as the foundation for a larger portfolio of responsive systems. To this end, we chose a radio transceiver that supports multiple radio protocols for easy integration into different network architectures. Our system is designed to support meshing with similar devices over Google's Openthread stack, in which each node has enough on-board resources to serve as an active mesh intermediary in addition to the resources required for simple edge computing. Mesh protocols are advantageous to future Internet-of-Things (IoT) adoption because they sidestep support and troubleshooting related to individual router infrastructure.

4.2 Extensibility

It was a major goal to enable other researchers to use Captivates for their own experiments without redesign work. We made the glasses easily reprogrammable without disassembly or rework, and fully open source, so anyone may utilize the device. We also included enough additional connectivity to simplify integration with outside sensing platforms, including support for multiple popular network stacks like Bluetooth.

Table 2. System Specifications and Comparison

Design Target	Captivates	Google Glass Enterprise Edition 2	Focals by North	J!NS MEME ES
Weight	44g	46g	72.57g	36g
Battery Life	10 hours	<8 hours (depends on usage)	18 hours (intermittent use)	12 hours
Battery Recharge Time	3 hours	1 hour (fast charge)	2 hours	2 hours
Ruggedization	None	Water and dust resistant	Water and dust resistant	None
Sensing Modalities	•Nose and Temple Temperature •Touch •Blink Rate •9-axis IMU •3D Location	•Microphone •Touch •Camera •Gyroscope •Accelerometer •Magnetometer	•Microphone •Ambient Light •Gyroscope •Accelerometer •Magnetometer	•Accelerometer •Gyroscope
Actuators	•Eye-facing LEDs •Externally-facing LEDs	•Display Module •Mono Speaker, USB audio, BT audio	•Display Module •Mono Speaker	None
Price	$200	$1000	$599	$420

4.3 Design for User Comfort and Signal Robustness

Since this system is intended to be worn throughout the day, across static and dynamic activities (e.g., sitting, walking, etc.), our sensing modalities were optimized for comfort as well as consistency across activity and environment. As a result, we avoided sensors that require significant contact force (uncomfortable for long term use) or skin adhesion (contact changes over time). Furthermore, electrode-based techniques that require this type of coupling often have reduced signal integrity in long-term, mobile applications due to the dynamic mechanical stresses at the electrode-to-skin bond during movement, and natural variation of other confounding electrical sources, i.e. cardiac activity, ocular movements, eye blinks and muscular activity [23, 25, 34, 61, 73].

4.4 Manufacturability

After considering the appearance and comfort of the glasses, we optimized the system for easy production and assembly. The design balances the part count and overall assembly time with ease of modification and extensibility. Our goal was to have a system that is geared toward mass production, allowing us to cheaply scale production of plastic parts and assembled printed circuit boards at a cheap cost. In addition, we also wanted the design to be 3D-printable to allow users to easily produce small batches with custom modifications.

5 SYSTEM DESIGN

In this section, we discuss the technical system design. Our goal is two-fold: first, to highlight engineering work beyond typical prototype design that went into making these more product-like, in service of of robust real-world operation and design. Secondly, we intend to highlight the structure of the technical system for researchers who might want to modify it. The design is open-sourced and available on our website at https://captivate.media.mit.edu/resources.html.

We divide the system into three logical parts: the electrical, mechanical, and firmware. A summary and comparison of our final system with other popular smart eyeglass systems is shown in Table 2.

5.1 Sensor Selection

Based on the design considerations outlined in Section 4, we cross-referenced the recent literature for physiological indicators of cognitive load and attention. Much of the recent research is focused on frontal lobe electroencephalogram (EEG). There are several initiatives already exploring this technique, including BrainCo [3], Muse [4], and AttentivU [46]. Additionally, as described in Section 4.3, electrode-based techniques aren't well suited for long-term use and during mobile activities. For these reasons, we decided to refrain from this modality.

Furthermore, contact-based sensing of this type always requires a physical, head-mounted wearable. One of our criteria for sensor selection was the ability to approximate our measurements with off-body techniques. Any insights from this research can thus be generalized (in specific environments) without the use of smartglasses per se– unfortunately, this kind of contact-less physiological sensing infrastructure is too cumbersome and expensive to scale across the many real-world environments visited by even a single user. Wearables present the best opportunity to rapidly scale data collection across users and contexts.

We converged on a non-contact temperature sensor (i.e., thermopile) on the nose, as nose temperature correlates with cognitive load [11]. However, since nose temperature is also influenced by external stimuli (e.g., air flow), a baseline temperature measurement is required. For this we added a second non-contact temperature sensor at the temple, which should not see rapid fluctuations during internal state changes [55]. The second sensor we chose was an eye blink sensor, as cognitive load influences blink rate [49, 52]. To accurately measure blinks, we are using an IR emitter/receiver pair since it is relatively low-power compared to cameras and can detect blink rate and intensity [27]. We also include an IMU, which allows us to capture a person's activity level and head pose [41]. Head pose is correlated with gaze direction, which is important to determine what a user is fixated under controlled conditions [67]. Lastly, we included VIVE Tracking System compatibility, so that off-the-shelf hardware can provide each pair of glasses a 3D location estimate, ideal for events with crowds and/or monitoring individual movement in a room. This can also be used for tracking fixations based on head pose and 3D localization data. Finally, the system also includes on-board user-facing LEDs that can subtly alert the wearer to any notifications (e.g., take a break, alertness is dropping, etc.) and influence their attention, as well as externally-facing LEDs for creative use in performance.

5.2 Electrical System

5.2.1 Circuit Layout. In order to maximize the amount of circuitry that can fit into a traditional eyeglass form factor, we spread the electronics across all faces of the glasses. This requires a three PCB-assembly design, one on either side and a flexible one running across the front that connects both side electronic assemblies. The side PCBs are standard FR4-core, at 0.8mm thickness to minimize the overall temple arm thickness. The front flexible circuit is the most novel piece of circuitry that we designed, as it conforms to the contours of the eyeglass's front plastic housing, folds down into the nose piece, and robustly handles dynamic bending within the hinge. We started with a 3D rendering of the flex that had the correct curvatures (Figure 2), which we then virtually unbent using sheet metal tools in our computer-aided design (CAD) software to achieve a flat footprint (Figure 3).

One of the most challenging issues to overcome with this system was how to incorporate the sensors around the nose into the electrical and mechanical design. From a design perspective, the easiest approach is to solder the near-nose sensors to the flex circuitry through a set of wires, but this method significantly lengthens and complicates the assembly process. Instead, we matched the bends of the flex circuit to that of the mechanical housing around the nose, creating two tails of flex circuitry that included the temperature and blink sensors. During installation, these tails fold into the grooves of the plastic housing without any additional soldering. The final manufactured flexible circuit is shown in Figure 3.

5.2.2 Processors. A block diagram of the entire system is shown in Figure 4. For the main processor, we use a dual-core microcontroller (MCU) that has an ARM Cortex-M4 processor for the user application and an ARM Cortex M0+ processor for the network stack. This processor allows us to treat the MCU as an application processor, without concern for network stack overhead. It supports quick embedded firmware development and increased versatility for applications that don't require the network component. It also supports three major network stacks (i.e., Bluetooth 5, Zigbee 3.0, and OpenThread). The chip is thus a versatile choice, minimizing the need for hardware redesign in a clean and simple electrical design that integrates this functionality into a single discrete package.

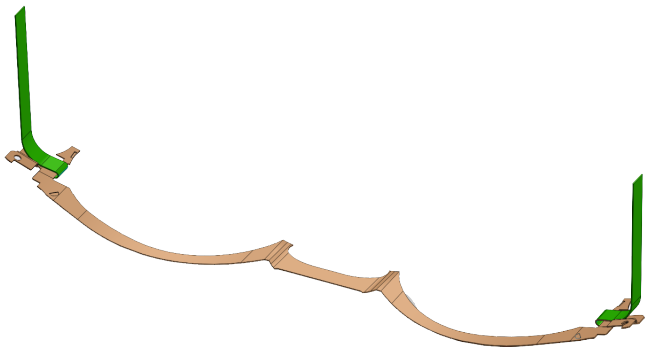

Fig. 2. Curved Flex Circuit Rendering

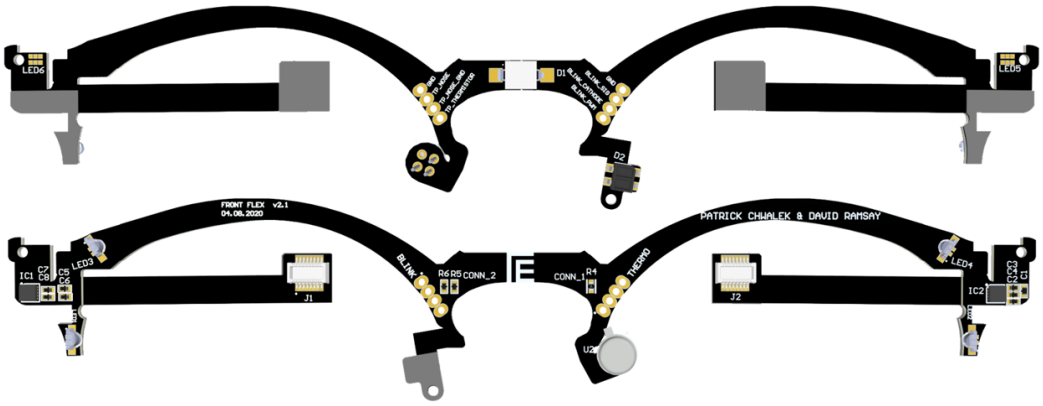

Fig. 3. Design of Flexible Circuit Board (Top is Front-facing, Bottom is Rear)

The dual-core MCU exists on one of the two-sided PCBs, which are separated and connected by the flexible circuitry. Given that sensors exist throughout the glasses, and that the long separation between the rigid PCBs can lead to signal integrity issues, we added a secondary Cortex-M4 MCU on the opposite side of the glasses. This secondary MCU manages the sensors nearby, applies our preprocessing algorithms to those sensor streams, compresses the data, and sends it to the main processor when ready or requested. While this increased the software complexity, it allowed us to reduce the number of conductors passing through the hinges and across the length of the flexible circuit.

5.2.3 Power Architecture. Our system utilizes two batteries for better weight distribution, but this approach increases the complexity of our power architecture and charging circuitry. As shown in Figure 5, the lithium-polymer batteries and their charge controllers are electrically separated in our system until they are interfaced with our power multiplexer (MUX). This MUX is required because it is unsafe to put these batteries in parallel for charging through one charge controller. The batteries are physically far apart (i.e., one in each arm) and

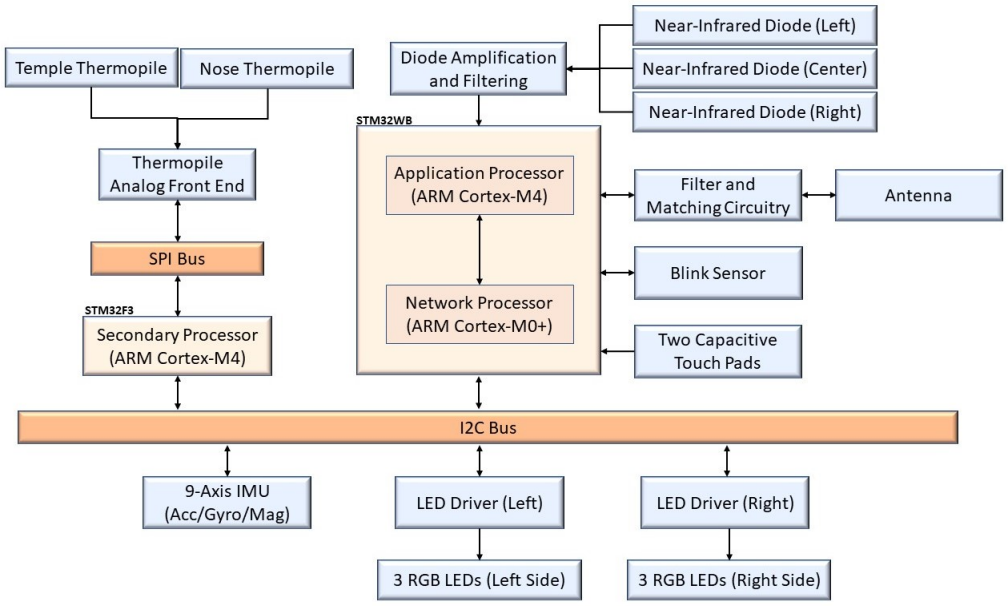

Fig. 4. System Diagram

experience different environmental conditions; additionally, batteries from different manufacturing batches may age at different rates. These conditions lead to possible cross charging (i.e., one battery depletes into the other) and rapid aging. This choice means we do not have to design impedance matched charge lines to each of the spatially separated batteries, which otherwise would have been difficult to model given the complexity of the multiple circuit designs. Furthermore, The MUX supports asymmetrical battery topologies, in case in the future we need to re-balance weight or add additional battery power to one side.

Within the multiplexer, the battery with the highest charge is prioritized until it discharges to around 0.2V under the idle battery. At that point, the comparator of the MUX toggles a switch and the idle battery is set to active while the active is made idle. If both inputs to the MUX are under 2.8V, the MUX goes into high impedance mode and shuts off the battery inputs and system output completely. This ensures that the batteries aren't over depleted, which can result in damage to the batteries and other hazards. The MUX output then powers two 3.3V, low-quiescent, linear regulators, one on each side of the system, and also directly powers the two LED drivers located on the flexible circuit.

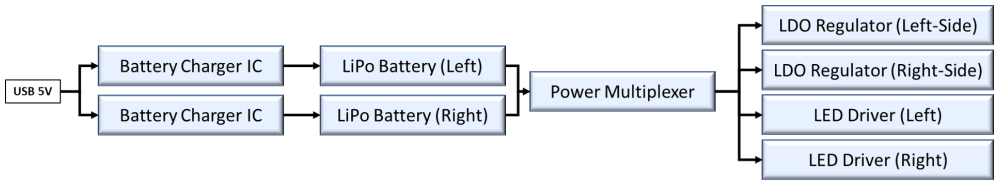

Fig. 5. Power Architecture

Fig. 6. Nose Thermopile

Fig. 7. Temple Thermopile

5.2.4 Temperature. The thermopile circuit consists of two analog thermopiles, one faced towards one side of the user's nose (Figure 6) and the other faced towards the user's temple (Figure 7). These thermopiles are then connected into an analog front end, which amplifies and filters the signal before being converted by the secondary processor's 12-bit analog ADC. Once a set of readings is collected by the secondary processor, the main processor is alerted over an interrupt line, and after acknowledgement, the data is transmitted over the I2C bus to the main processor.

5.2.5 Blink. The blink sensor is an IR emitter/receiver package that showed promising results in [27] on detecting blink events. We experimented with different blink detection locations and found that putting the sensor on the bridge of the nose, facing towards the eye, yielded the best performance. The near-infrared diode that emits light can be modulated through the main application processor to better dynamically adjust the reflected power seen by the receiving diode and enable synchronous detection. This feature is important for environments that have external near-infrared light sources (e.g., sunlight) that can saturate our blink sensor.

5.2.6 3D Localization and Pose Estimation. For the 3D localization, we are using two VIVE base stations [1] as our signaling sources in the environment in order to triangulate the glasses in 3D space. Each base station emits a vertical and horizontal near-infrared sweep, alternating between both sequences and between both base stations. Prior to each sweep, a pair of near-infrared LED flashes occur that signals to the receiver (i.e., the glasses) which base station is sweeping and if the sweep is horizontal or vertical. Ultimately, a set of four sequential sweeps (i.e., 1 horizontal and 1 vertical from each base station) repeats at a rate of 30Hz and is received by a near-infrared receiver. With the receiver being calibrated to know the global coordinates of each base station, it can determine where it is in 3D space by identifying the sweep and comparing the timing between sweeps. VIVE states that a less-than 2mm accuracy and a range of 16 feet can be achieved with their system, but based on preliminary testing we believe our tuned custom system can detect the base station sweeps at a range of at least 30 feet.

The hardware architecture for our receiver consists of three near-IR sensitive diodes, one on each face of the glasses. The search criteria for choosing which type of diode to use consisted of finding a surface mount component that had a slim form factor but a large enough receptive area that also offered adequate responsivity to capture the momentary sweep. In addition, the field-of-view (FOV) needed to be large enough as to limit any blind spots. We evaluated six types of diodes and found that the VBPW34FASR by Vishay performed the best and conformed to our design constraints. The three Vishay diodes in our circuit are then tied together (i.e., summed) and passed into a transimpedance amplifier to convert the diode-generated current into a voltage. The reason for the summation is that we wanted to treat the system as a point mass that has a large FOV so that as a

person turns their head, at least one diode will still be exposed, increasing the reliability of the system. This also allows for a simpler analog front end with the trade-off of decreased accuracy, since the diodes are not spatially co-located but are treated as such in software.

For head pose, we chose a digital 9-axis IMU that connects to our I2C bus and offers various internally computed metrics, such as activity classification and a quaternion matrix for estimating head pose. Software calibration was done to align the IMU's reference coordinate system to the glasses coordinate system for accurate head pose estimation.

5.2.7 Other Components and Interactivity. Our system offers a few other components for easy usability, interactions, and actuation. In order to easily program our system, we wired the USB micro port to both the battery charging circuitry and to the single-wire programming interface for the main application processor through a modified USB cable and an ST-Link Programmer— this can be done by anyone with just a programmer and a spliced USB cable. In addition, the system has two capacitive touch points on the left arm that are sensitive enough to identify taps and both forward and backward swipe gestures. The sensitivity of the touch points can be adjusted in software to tune for specific applications. Finally, as shown in Figure 8, there are six RGB LEDs, one facing towards either eye, two on the top, and two facing forward. The LEDs toward the eye are intended to be used as notification [22] while the other LEDs can be used for crowd applications (e.g., a concert where the LEDs are actuated by the person's physiological signals). Our electrical design features several GPIO breakout pads scattered throughout so to make retrofitting for other applications easier.

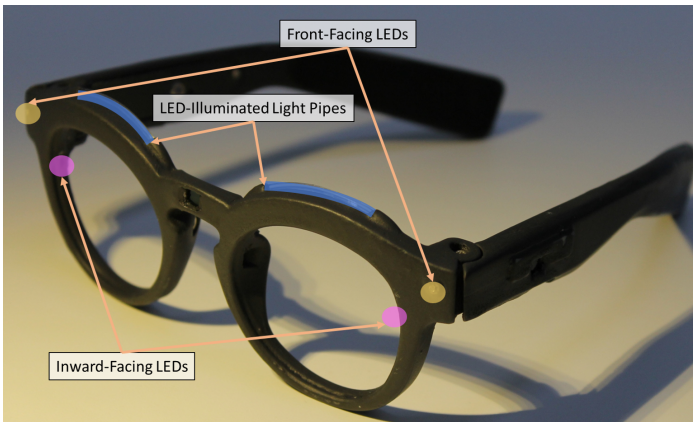

Fig. 8. LED Locations

5.3 Mechanical Design and Assembly

The mechanical design of the glasses required several iterations as we converged to a final form factor. As outlined in Section 3.1.3, we wanted to choose a design that looks less like a research prototype and more like a traditional pair of glasses that is acceptable to wear in public contexts. Figure 1 shows the front view of our most recent design. Openings above the bridge of the nose and on either side of the glasses expose the near-IR diodes to the incident light emitting from the VIVE base stations. These openings may be covered with near-IR transparent plastic, but we've left them open for easy debugging. These openings can be covered or completely omitted for applications that do not require the 3D localization feature.

Fig. 9. Pre-bent Flex Circuitry

Fig. 10. Inside of Front Housing

Fig. 11. Sealed Front Housing with Flex Circuit Integrated

Fig. 12. Adding a Side Leg

5.3.1 Hinged Design. We iterated on various hinge types, including initial attempts where the hinge was omitted entirely and the glasses were rigid. The rigid option proved to be nonideal since it made it difficult to transport the device between locations due to the fragility in the rigid arms acting as cantilevers on the joint. This early prototype was also seen as being less socially comfortable to wear when we polled our peers and later validated it in our comfort study (Section 6.2) so we know a redesign was warranted if we wanted long-term, daily use. As we talked with more glasses designers, we learned the importance of designing spring action into the glasses. Usually these spring points are at the hinge and above the bridge of the nose, and flexible plastics like Nylons are used. After a rigid prototype, we attempted to use off-the-shelf spring hinges for glasses as a starting point. Eventually, we settled on an approach that is commonly found in smart eyeglasses, designing a completely custom plastic hinge with the flex circuitry running through the middle of it. Our flex circuitry is preformed (Figure 9) and inserted into the front housing, resting on internal raised plastic features (Figure 10). The flex circuit is then sealed by a piece of plastic acting as the back plane (Figure 11) and then the arm is able to be rotated around the flex circuit (Figure 12) and inserted into the plastic via a fastener running through the rotating axis of the hinge. The hinge preserves a wide bending radius in the flex PCB in open and closed positions, as well as plastic guards to protect it from direct exposure to the external environment.

5.3.2 Light Pipes. One of the design features we were most excited about was to have the glasses include a feature that allowed them to transition from normal day use to a more festive environment where we can visually actuate individuals and groups based on internal and external stimuli. To do this, we included a few RGB LEDs, including one on each brow that would light up the entire top ridge of the glasses. One way we attempted to do this was by crafting a piece of polymethyl methacrylate (PMMA), also known as acrylic, to form a light pipe for the light from the RGB diode to transmit through. For our design, we needed the light to diffuse through the acrylic and evenly illuminate only the outward face of the light guide while internally reflecting on all the others for maximum efficiency. We worked with a manufacturing facility in China to laser cut these PMMA pieces and chrome-coat three of the four sides to maximize the light emmission on the top-face (Figure 13). Although this method worked, we found it to be cost prohibitive given that a single piece cost us over $50 to custom manufacture.

An alternative approach that we settled on was to use side-glow fiber optic cables. A traditional fiber optic cable is not suitable, since it lacks additional cladding on the cable's surface that allows for greater diffusion of light across the length of the cable. Side-glow fiber optic cables, however, have a cladding, usually clear Teflon, that decreases the critical angle of incident on the internal surface that is required for external emission of light. Figure 14 shows the fitted 1.5mm diameter side-glow fiber optic cable illuminated in our design. Apart from the brow lighting, we also have lighting at two points on the front face of the glasses where some traditional glasses

Fig. 13. PMMA Light Pipe: Chrome-Plated (Top) and Non-Plated (Bottom)

Fig. 14. Illuminated Brow of Glasses

Fig. 15. One Sided Commercial Battery Design of Smart Eye-Wear

Fig. 16. Battery Placement in Captivate's Arm

have rivets. To increase light visibility, we added commercial-off-the-shelf PMMA light pipes with fresnel lensing on one end so that the emitted light is less directed, offering better angular visibility.

5.3.3 Battery Placement. A tip we received from the manufacturers in Shenzhen is that battery placement in the glasses is important for weight balance and power capacity considerations. Batteries are usually the most dense component in a pair of smart eyeglasses so their positioning matters due to weight distribution. If placed too far forward, the glasses can slip off easily and if placed behind the ear, can lead to a less traditional aesthetic due to the battery's large form factor. There are batteries with curved designs that can potentially go around the ear but those designs come at a trade-off of decreased power density and likely require custom manufacturing. Oftentimes, designers put a single battery on only one of the glasses (Figure 15) to simplify the design, but any imbalance can lead to concentrated stress regions on the nose pad due to the battery's mechanical moment. The ideal method is two identical batteries, one on either side, to better balance the weight at the cost of complexity in power distribution architecture and charging circuitry.

For our design, we have two, 150mAh, batteries on either side of the glasses, running in series with the side printed circuit board. The circuit boards are designed so that when assembled, the width is uniform across the batteries and across the circuit board assemblies for a streamlined design (Figure 16).

5.4 Firmware Architecture

For a system of this complexity, given all the sensor channels, potential future channels, and preprocessing steps, it's warranted to use a low-level operating system to help keep track of all the events, prioritize any events that need more real-time action, and provide the utility for easy debugging. To do this, we decided on using a real-time operating system that already has native support for the processors we are using and has a variety of tools that aid in debugging and in modeling the system's performance.

5.4.1 FreeRTOS. We decided to use an open-source real-time operating system for this project since it simplified the software design and makes it easy for new researchers to disable current software subsystems or add-on new functionalities. FreeRTOS is a lightweight real-time operating system that is capable of being run on microcontrollers and small microprocessors which makes it suitable for this project [7]. This operating system is also supported by ST Microelectronics, the manufacturer of our processors, and has several additional tools written by ST Micro for easier system integration. FreeRTOS allows for the encapsulation of individual blocks of code into threads that can be triggered by a variety of sources (e.g., hardware interrupts, timers, other threads, etc.) and has various methods of passing messages across threads (e.g., queues, semaphores, mutexes, etc.). This encapsulation allows you to think of the firmware system as embedded blocks of code that have a deterministic operating behavior. In addition, FreeRTOS offers various debugging tools that allow you to assess the memory and time utilization of the processor in real-time to model the system and determine if certain threads are causing delays or deadlocks. This allows for more efficient debugging since the problem can be quickly pinpointed to specific logical blocks.

As shown in Figure 17, there are a variety of threads in our system, all controlled by a master thread within each processor. The master thread receives queues from either the radio communication thread or inter-processor thread to enable certain sensor channels, actuators, or to send or receive messages. After enabling, each sensor channel is triggered by the hardware resource it uses for the most efficient operation. For the blink sensing, the direct memory access (DMA) controller is used so that when sampling the blink sensor at 1kHz, the thread only awakens every second (i.e., 1000 samples) to preprocess and pass the data pointer to the master thread where the data is packaged for transmission. For inertial sensing, the digital IMU interrupts the processor when data is ready to be transmitted, triggering the Inertial Sensing Thread to start I2C transmission. The secondary processor also interrupts the main processor once it buffers 10 samples from each thermopile sampled at 10Hz and waits for an acknowledgment before sending the data over I2C.

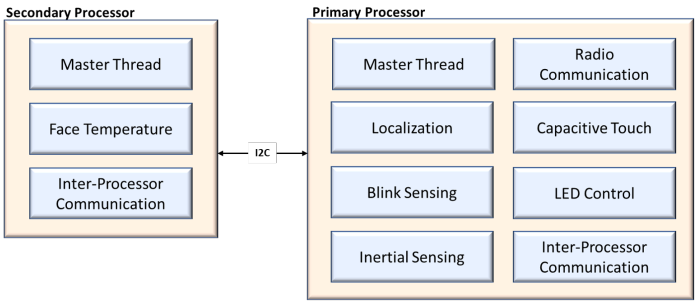

Fig. 17. FreeRTOS Threads in System

5.4.2 3D Localization. The 3D localization code was adapted from prior work done in [65]. Every time one of the near-IR diodes is saturated by incidence light (i.e. rising or falling edge), an interrupt is triggered which results in

a callback recording the exact timestamp of the event in microseconds. These timestamps are then fed into a thread that classifies if the event was part of a sequence (i.e., two flashes and a sweep) or if the event is noise. If part of a sequence, the code catalogs all the sweeps and calculates an estimate for the 3D location of the device. Since the code requires the global coordinates of the VIVE base stations, they can either be hardcoded into the software or updated via the server.

5.4.3 Networking: OpenThread. For networking, there are a few protocols that we could have chosen to use with our network processor. In the end, we wanted to use something that was scalable, supported across a variety of platforms, and could integrate well with server infrastructure and ultimately, the internet. The most ideal candidate was an open-source version of Thread [8] called OpenThread [6]. It "is an IPv6-based networking protocol designed for low-power Internet of Things devices in an IEEE 802.15.4-2006 wireless mesh network, commonly called a Wireless Personal Area Network (WPAN). Thread is independent of other 802.15.4 mesh networking protocols, such a ZigBee, Z-Wave, and Bluetooth LE" [6]. OpenThread has a variety of features that makes it suitable for our system. One primary advantage is that it's a protocol that is seeing wide adoption across a variety of platforms, and since its being supported by Google, it's integrated into many of their home devices—in theory, our system could use Google Home nodes as hops in our network. It's also scalable to hundreds of nodes, making it ideal for future integration of our glasses into an ecosystem of sensors and actuators. In addition, it's IPv6-based so integrating it with other IPv6-based systems is simple and doesn't require a network translator, making it appealing for a "future-proof" implementation. There also exists Contiki [30], another low-power IPV6-based protocol with evidence that it has improvements in latency and packet loss rates over OpenThread [31]. However, [31] was just an introductory comparison between both protocols, and more work needs to be done on measuring the power efficiency and performance of both protocols. We chose to stick with OpenThread due to the native support ST offers for our network processor and the active open source community around it.

5.5 Network Architecture

In this section, we describe our work on the network infrastructure that allows the system to easily scale for multi-user applications. We also present a dashboard we designed to allow for researchers to monitor and interact with all the connected active systems.

5.5.1 Server. OpenThread has a variety of node types that can exist in a given network [6]. One is a Full Thread Device (FTD) that is classified as a node that is always on and can be used as one of the 32 main routing nodes in the network. Another type is a Minimal Thread Device (MTD), which is a node that always needs a parent to forward traffic through and can go into a sleep mode to conserve power. Within FTDs, there exists a role type of Border Router that is a device that "can forward information between a Thread network and a non-Thread network (for example, Wi-Fi). It also configures a Thread network for external connectivity" [6]. This type of role is suitable for a server, since it has direct access to the thread network and can connect to the internet to outsource the data. Therefore, for indoor applications with internet accessibility, we use a Raspberry Pi equipped with a network node that is configured with the OpenThread Border Router (OTBR) build of the network stack to interact with our system through any internet-connected computer using a dashboard that will be described in Section 5.5.3.

5.5.2 Server Protocol. For the actual server component of the network, our OTBR uses a custom implementation of the Constrained Application Protocol (CoAP) [64]. CoAP is a "specialized web transfer protocol for use with constrained nodes and constrained (e.g., low-power, lossy) networks" [64]. CoAP is ideal because it is "designed to easily interface with HTTP for integration with the Web while meeting specialized requirements such as multicast support, very low overhead, and simplicity for constrained environments" [64]. There exist a few open

source implementations, so we chose to modify a build of CoAPthon [70] since it was written in Python and was relatively simple to add multicast functionality, a feature that isn't often fully supported in other builds of CoAP.

Each eyeglass node in our network hosts a few server endpoints that other nodes can query to retrieve the node information, such as unique indentifier (UID), node type, sensor status, and server time. Each node can also have its real time clock synchronized by the OTBR and have other settings modified as well (i.e., enable/disable sensor channels, actuate lights, and unicast to OTBR). The OTBR's CoAP server is the main endpoint to where all the eyeglass nodes send their data when logging for an experiment. However, it's easily capable for one eyeglass node to send another few metrics that can be used to synchronize actuations. One example that can be implemented is synchronizing lighting across glasses based on collective physiology measurements to visually identify how an audience is reacting to a performance.

5.5.3 Captivate's Dashboard. The OTBR has also been configured to send and receive encrypted packets over the internet to a graphical user interface (GUI) dashboard that we designed for our system (Figure 18). The dashboard was built using PyQt and offers various control and visualization tools [56]. First, it allows you to connect to one of our OTBR nodes and view a list of all the nodes in the network with their IP, node type, short description, and unique identifier shown. The dashboard allows you to select which sensor channels to enable on the glasses and to toggle streaming of those sensor channels. On the right, you can visualize the signals being transmitted from the system in real-time. On the left, you can see the live quaternion values, a list of activities and how confident the system's classification is of each activity being performed (e.g., walking, running, sitting, etc.), and their approximate 3D location in a room if the VIVE system is active. If the inertial sensor channel is activated, there is a secondary window that appears that shows a block rotating in the same directions as the glasses. The dashboard also allows you to manually toggle and set the color for each LED on a pair of glasses and offers the ability to orient the inertial system in order to properly calibrate the IMU in the reference frame of the glasses.

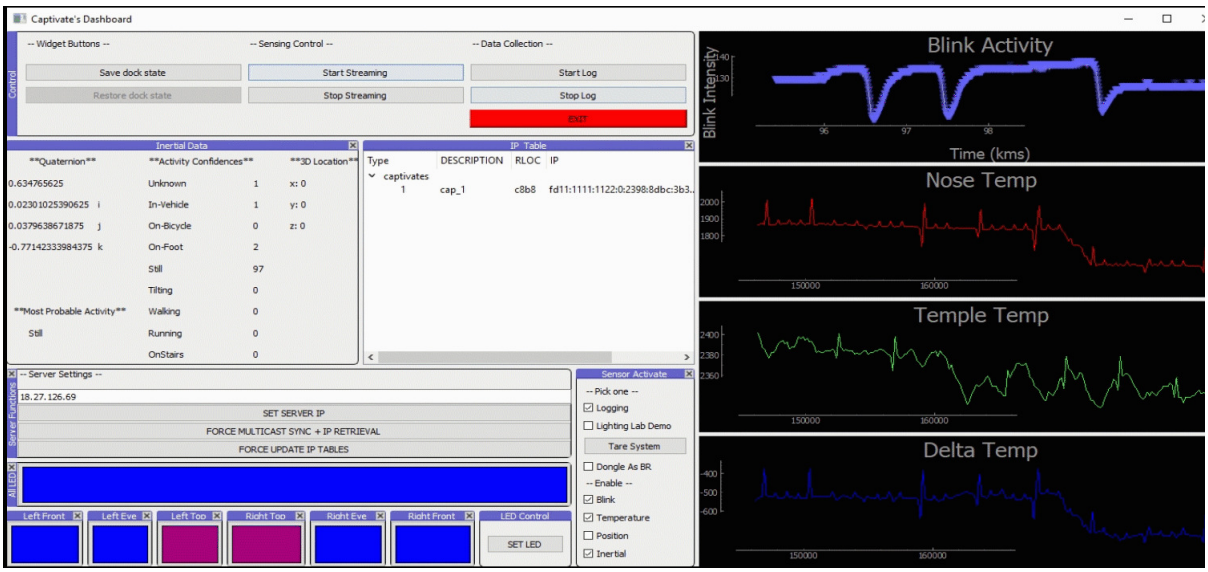

Fig. 18. Graphical User Interface for our OpenThread Border Router

5.6 Production and Extensibility

We opted to manufacture our plastics using Nylon PA12 through a selective laser sintering (SLS) 3D printing process using an external vendor since Nylon is much more resilient and robust than other 3D printing materials while SLS printing offers a more uniform surface finish. However, we were also able to print all of our mechanical components with consumer-grade 3D printers (e.g., Formlabs Form 2, Prusa) so that is an avenue that is available for quick experimentation.

Our two rigid circuit boards were designed with specifications that fit most common circuit board manufacturers' capabilities (i.e., 4/4mil traces, 4 layers, 0.8mm board thickness, 0.2mm minimum hold sizes, and 1oz copper). The front flexible PCB is the only design feature that has a custom design feature, specifically having transition regions between 2-layers and 4-layers. We worked closely with a manufacturer in Shenzhen on producing the boards so it is possible other manufacturers may be able to support this feature. Otherwise, you can avoid the transition altogether and just make a 4-layer flexible PCB which would trade-off robustness for added simplicity. For circuit board assembly, we were able to assemble a few sets by hand in our lab but many of the components are small and the front flexible PCB has 2 ball grid array (BGA) components so if lacking the soldering capabilities and expertise, an outside assembly service is recommended.

For researchers wanting to add additional sensing modalities, we've added a few thru-hole connections on the PCB located on the right side that provides 3.3V power, ground, and connections to 3 general-purpose input/output pins on the secondary MCU, equipped with access to the I2C communication and analog-to-digital conversion (ADC) peripherals. If some of the existing sensing subsystems are not needed for an application, those sensors can be removed and the existing connections can be re-purposed. One example would be to remove the blink sensor and use the existing power and the ADC lines to sample the signal of a different sensor. To make this easier, we've added various thru-hole regions on the front flexible circuit that wires can be soldered to. Another method that can be used is removing the PCB with the secondary processor and reusing our existing power circuit to design a new board with other sensing modalities.

All the electrical and mechanical design files are linked on our website, as well as further information on our progress with the system: https://captivate.media.mit.edu/.

6 VALIDATION AND DISCUSSION

6.1 System

We calibrated the antenna circuit using a vector network analyzer (VNA) while the PCBs were in the eyeglass frames and on a users head. This approach is warranted since any mass near the antenna can attenuate the signal and distort the balance of the circuit. We also calibrated our 3D Localization System by tuning our circuit to barely saturate at 30-feet from the VIVE emitter, ensuring our signal-to-noise ratio is optimal for that range. In the next subsection, we describe our detailed calibration of the temperature sensors.

6.1.1 Temperature. To estimate cognitive load, you only need to measure relative temperature differences from the baseline [11] so thermopile calibration for absolute temperature approximation is less important. However, for applications that require an accurate estimation of absolute temperature, a thorough calibration should be performed on each individual thermopile since different resistive losses and noise artifacts exist based on the location of the sensors.

As shown in Figure 19, a thermopile is a series of thermocouples with one set of junctions fixed to the sensor housing (reference junction) while the other set (active junction) is coated with an infrared absorbent material and exposed to ambient light, usually blocked by a far-infrared transmissive material to filter the incident light. As a temperature differential forms between the two sets of junctions, a voltage differential is created through a process known as the Seebeck Effect.

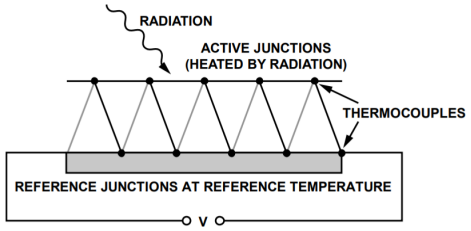

Fig. 19. Thermopile Architecture [15]

As explained in [36], the total radiation emitted by an object that does not selectively emit or absorb specific wavelengths is defined as

$$P_{obj} = \sigma\varepsilon(T_{object})^4 \tag{1}$$

where P_{obj} is the total radiation power emitted by an object, T_{object} is the temperature of the object, σ is the Stefan-Boltzmann constant, and ε is the emissivity of the object. For many objects, the emission factor is between 0.85 to 0.95, where 1 is for an ideal black body. Equation 1 is formally known as the Stefan-Boltzmann's Law, and combining it with the Seebeck effect, we get

$$V_{TP} = A(T^4_{object} - T^4_{reference}) \tag{2}$$

where V_{TP} is the thermopile voltage, T_{object} is the temperature at the active junction, and $T_{reference}$ is the temperature at the reference junction [43]. The constant, A, is a product of the thermal resistance of the thermopile, number of thermocouples within the device, the Seebeck coefficient, net emissivity between the object and the device, Stefan-Botlzmann constant, and field of view (FOV) of the device. In most cases, this constant is approximated for during calibration by varying the temperature of an object and of the thermopile housing, while placing calibrated thermocouples throughout. Rearranging equation 2 to get T_{object} as a function of V_{TP}, we have

$$T_{object} = \sqrt[4]{T^4_{reference} + V_{TP}/A} \tag{3}$$

However, in [71], it is recommended that we also account for heat flow other than radiation, such as convection and conduction from nearby objects. We modify our equation as such,

$$T_{object} = \sqrt[4]{T^4_{reference} + f\{V_{TP}\}/A} \tag{4}$$

where

$$f\{V_{TP}\} = (V_{TP} - a_0) + a_1(V_{TP} - a_0))^2 \tag{5}$$

and a_0 and a_1 are terms to solve for during the calibration process.

For our system, to estimate each thermopile's reference temperature, $T_{reference}$, each thermopile has a thermistor built-in with a manufacturer-provided characterization [14]. To calibrate, we took a Peltier module that converts a voltage potential into a temperature differential and using a thermal compound, embedded a calibrated thermocouple between the Peltier module and a thin piece of laminate material. We then placed a calibrated digital thermopile right next to our analog thermopile as a reference. Finally, using thermal compound again, we attached a thermocouple to the housing of the thermopile to estimate the accuracy of the built-in thermistor. A diagram of the test is shown in Figure 20 with a snapshot of the experiment in Figure 21.

The test was performed over 10 different temperatures for both the nose and temple thermopiles, ranging from 60-115 degrees Fahrenheit, the average temperature ranges expected across the surface of a person's skin. For each iteration, we waited a minimum of 5 minutes to ensure the temperature of the reference object has stabilized. After the experiment, we used non-linear least squares to solve for the constant A given V_{TP}, $T_{reference}$,

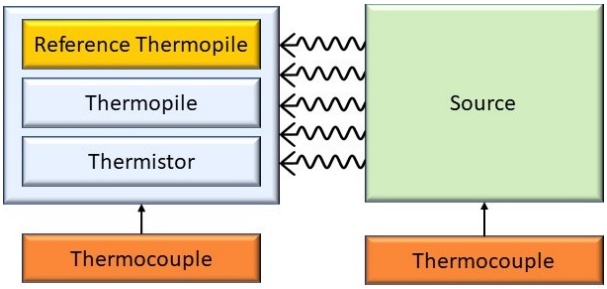

Fig. 20. Thermopile Calibration Test Setup Diagram

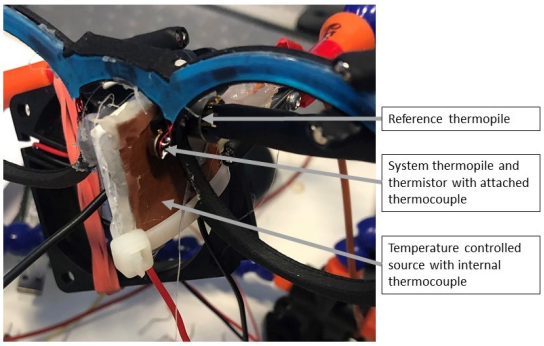

Fig. 21. Thermopile Nose Calibration Test Setup

Table 3. Thermopile Calibration Values

	A	a_0	a_1
Nose Thermopile	7.80e-10	-2.31e-01	3.61e-03
Temple Thermopile	4.21e-10	-3.62e-01	8.31e-02

and T_{object}. This allows for a suitable approximation for temperature within our ranges but assumes constant emissivity. This assumption doesn't hold true given the variety of skin types, so a more thorough calibration is warranted where the emissivity of the object and $T_{reference}$ are varied. Table 3 shows our results from the least-squares approximation; these values compensate for differences in resistive loading for each thermopile (e.g., variable trace length), in addition to calibrating the thermopiles for absolute temperature measurements.

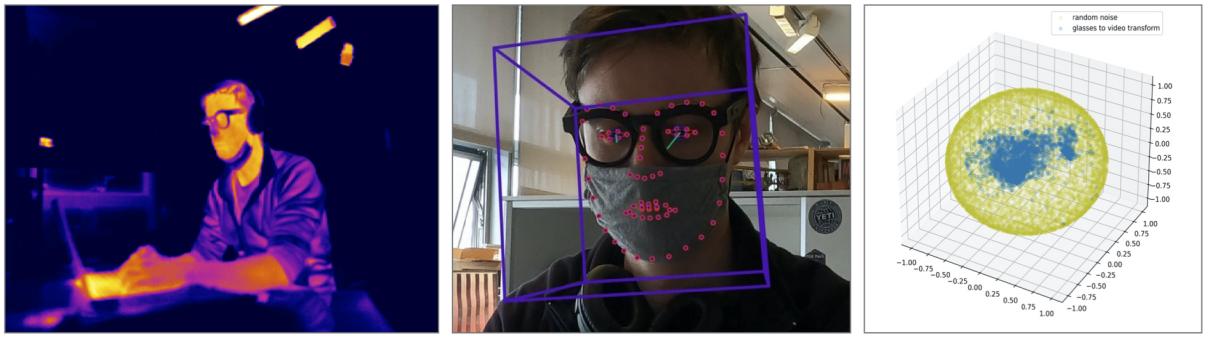

Fig. 22. Thermal imaging (right), the OpenFace video analysis tool (center), and a comparison of orientation data between the glasses and extracted from the video (a tighter blue cone is better).

6.1.2 Real World Evaluation. We validated the system in a pilot study in which 5 participants (2 female/3 male, ages 26-34) wore the glasses while being videotaped using a GoPro set to 60 FPS/1080P. Participants were recorded in naturalistic environments (their offices or homes) working as normal. Over 45 hours of data were collected in this way (>9 hours per participant on average).

This data has high ecological validity– participants have a range of face sizes and morphologies, studies were done at various points in the day from 10 AM to midnight in various lighting conditions and environments, and gaze and blink patterns varied dramatically over participants and over tasks. One participant made infrequent but clear shifts in visual attention between their screen and a desk; another participant shifted gaze frequently between a phone, a laptop, and a television screen; a third was making frequent large gaze shifts across a large monitor. Recorded activities were screen-based because we needed long-term, consistent and clear video recording of participants' faces. Blink behavior across participants varied dramatically, from an average of 11.3 blinks/min to 62.1 blinks/min over individual participants (31.0 blinks/min average across users).

We performed three basic comparisons to ensure our glasses data was of high quality. For two of these comparisons, we used the OpenFace machine learning classifier tool on our videos as a reference point (center of Figure 22) [19]. OpenFace uses trained machine learning models to calculate facial landmarks that are then used to estimate facial muscle movements (e.g., blink, smile, eyebrow raise, etc.). OpenFace struggled with our naturalistic videos– especially due to the framing and camera angles that we were forced to use when installing them in participants' real workspaces. We strived to install the camera in a way that the eyeglasses do not occlude the eye area of each participant from the camera. Due to the nature of how OpenFace calculates landmark positions based on existing models, it's uncertain to use if occluding part of a subject's face reduces the accuracy of the initial landmark estimation. For two participants, OpenFace's self-reported successful classification rate was between 25 and 40%. With manual cropping and editing, we were able to improve these classification rates to between 65 and 85% per participant. For all further comparisons, we utilized contiguous video sections longer than 200ms (the duration of a blink) where no more than 2 consecutive frames were dropped by the classifier. Time alignment between the video and glasses data was performed by looking for common minima in an extracted blink signal cross-correlation across several of these video sections. This time alignment was verified visually with a data-video overlay.

1. Thermal Comparison We did a few simple comparisons between an Optris PI400 infrared camera (382x288 pixels, .08°C sensitivity, and ±2°C accuracy), pointed at the side of the face where we are collecting thermopile data, to ensure proper functionality of our thermopiles in realistic settings (as seen in Figure 22L). This basic comparison revealed agreement over several times between the Optris and the glasses data, within the error of our Optis measurement. While our data is aligned, even these high-cost IR cameras do not have the pixel resolution to precisely capture 5mm^2 patch of skin underneath our thermopiles at a distance that still allows for naturalistic participant behavior (the entire nose is blurred over 20 pixels at the demonstrated resolution and distance, and this effect of this spatial averaging varies with subtle participant motion). We expect higher resolution, and more spatially consistent data from our thermopiles than an IR camera can provide; however, this did serve as a reasonable verification that the calibration process described above was functioning properly.

2. IMU Comparison IMU data was referenced against head orientation as extracted from OpenFace. OpenFace uses a world coordinate system at the frame rate of 60Hz, while our IMU data is referenced to the head and captured at 10Hz. To check for agreement, we interpolate the IMU data to our IMU timestamps and calculate the orientation transformation required to map the OpenFace orientation to the IMU orientation at that moment. Figure 22 (right) shows these transformations for each time-step over several minutes for Participant 5. Ideally, the mapping between the IMU and OpenFace data would consistently fall at a single point on the unit sphere (one reading maps exactly to another with the same coordinate frame transfer). If the IMU data and the OpenFace data were uncorrelated, we'd expect the yellow unit sphere pictured. Naturally, we get a cone centered around the coordinate transform, which shows the combined error in measurement between the IMU and the OpenFace classifier. The error we see in this cone has a standard deviation (in Euler angle °) of [20.4, 9.8, 22.2]. This demonstrates a rough agreement of motion from one measurement to another across time. Participants often move around their environment in the study, and the quality of the OpenFace orientation extraction as a face frequently translates forward and backward is un-characterized; however, this check combined with other basic

checks for drift leads us to believe the IMUs are functioning within the manufacturer specifications of <5 degrees of absolute error at all times, capturing transient behavior that is correlated with activity, physiological state, and mental state.

3. Blink Comparison To ensure blink data from the glasses was reliable, we designed a basic, micro-controller-friendly algorithm to extract blinks from the IR signal received by the glasses. This algorithm not only captures blink timestamps; it captures useful metadata about each blink that has been shown to correlate with alertness [12]. The algorithm works as follows, and was applied to each participant without personalization (See Figure 23 for representative data):

1. Low Pass Filter the raw 1kHz IR data (light orange) using a 50-point moving average filter (green).
2. Take the first derivative of this signal (red) and smooth it with a 25-point moving average filter (purple).
3. Using a fixed threshold of -0.05, wait for 100 peaks in the first derivative signal that dip below this. At this point, we predict the user is wearing the glasses.
4. Record the next 100 minima below -0.05. We know from preliminary testing that >10% of these are almost assuredly real blinks, and >10% are almost assuredly not, across all users and conditions.
5. Calculate the average value of the 10 shallowest peaks (a noise metric close to -0.05) and an average of the 10 largest peaks (known true blinks) from these minima; calculate a threshold value a quarter of the distance between them, just above the noise level.
6. Compare the running first derivative signal against the threshold; whenever it drops below this value, we consider a blink to have begun. For the next 175ms, we monitor the first derivative and capture the minimum eye close velocity (the negative peak), the maximum eye open velocity (the corresponding positive peak), and the difference in timestamp between these two (a blink duration). This search window and duration are shown when the blink detector triggers in a faint brown (search window) and dark brown (extracted duration).

Initially, we planned to use OpenFace [19], a common standard reference for face tracking, as our baseline for blinks. We quickly realized that OpenFace struggled to locate blinks properly within our data, even after hand-cropping the videos to increase classification success. Moreover, OpenFace's 'confidence' metric did not directly map to the blink signal– when participants gaze downward at a desk or a phone, they can appear almost as though their eyes are closed. This is a situation where the head remains confidently 'tracked' but blink detection is very difficult for the classifier.

To overcome this, we hand-labeled over 2 hours of blink footage, capturing 2586 blinks (>500 per participant). These labels are done by watching footage (in slow motion for fast blinkers) and tapping in time on a keyboard– as you can see from the blue outlines of hand annotations in Figure 23, they are not always perfectly accurate. To handle these timing errors, we use a wide margin for error of several hundred milliseconds when estimating the quality of OpenFace and our glasses data.

We hand-labeled only footage that had been pre-selected for its high confidence from OpenFace but we were still forced to skip about 15 min of footage that contained sections that were difficult for labeling (i.e., the head was at an angle such that blinking was occluded). Our results for OpenFace blink detection thus overstate the quality of OpenFace blink detection based on the provided confidence metric, since we've thrown away about 1/5 of footage that contained some of the most difficult and ambiguous 'confident' OpenFace blink data.

For our glasses, while the IR emitter is normally on to ensure a strong received signal, we designed the circuit to latch the emitter off in bright IR conditions to prevent diode saturation. We did not encounter this condition in our testing (though many of our indoor environments had bright ambient light from nearby windows); the received signal was consistently in a good envelope. We did, however, detect a couple of small instances of bursts of environmental IR pollution from other electronics. These instances are easy to identify because they cause

consistent and repeated blink detection at a constant rate over a short period. We removed an additional 2-3 minutes' worth of data from the two hours of our recording because of this issue.

Our results are listed in 4. The glasses algorithm outperforms the OpenFace classifier by a wide margin, even for a relatively constrained naturalistic test with a generic blink detection algorithm. Breakdowns in classification accuracy are quite different for both algorithms. OpenFace struggles when people are looking down (i.e. at their phone), which they do a lot in natural settings.

Ironically, participants with rich IR glasses data streams perform the worst with our blink detection (lower participant as compared to upper participant in Figure 23). This data captures more secondary eye movements– it is marked by good IR coverage of the eye, high contrast between pupil and cornea, and lots of attention shifts and saccadic movement– and these details can significantly alter reflected light, which engages the blink detector.

In the raw IR data, we can see characteristic signals of a blink, but we also see the opportunity to extract gaze/attention shifts (looking down at a phone or desk vs up at a screen) and even saccadic eye movement in some cases. With a more complex future algorithm that incorporates metadata about the blink, signal shape, and/or classifies other eye movements like gaze shifts, we should be able to do an even better job than this baseline technique.

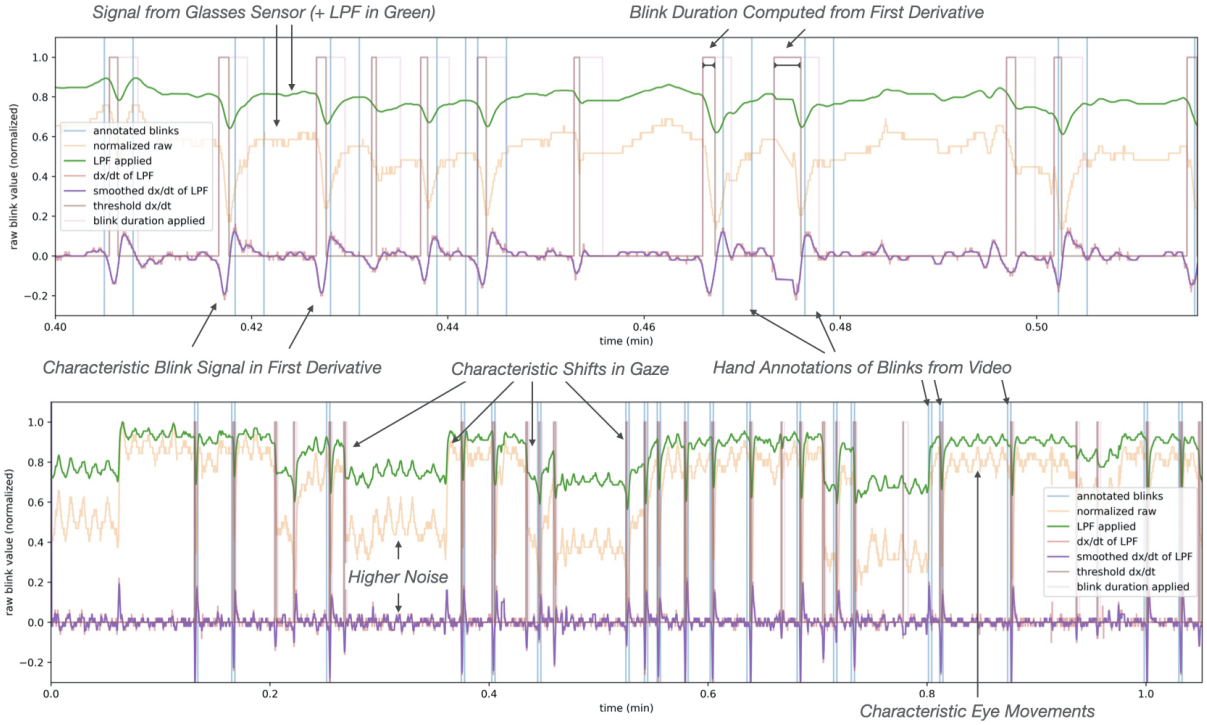

Fig. 23. Representative blink trace data from Participant 3 (bottom) and Participant 5 (top).

These early results have been further corroborated in a separate pilot user study with a further 6 participants, in which each user performed half an hour of cognitive tasks (IRB protocol #2001000083) alongside video recording. Initial results show comparable performance to off-body systems across subsystems and users in this test as well.

These results give high confidence along a few major dimensions: first, our sensor geometry is designed well enough to accommodate a diversity of head sizes and eye/nose angles. Secondly, we see useful data capture

Table 4. OpenFace and Glasses Algorithm Success for Detecting Blinks Across Participants.

	OpenFace		**glasses algorithm**	
participant #	*sensitivity (%)*	*false discovery (%)*	*sensitivity (%)*	*false discovery (%)*
1	38.4	21.7	91.9	5.0
2	53.3	58.0	88.7	33.8
3	67.7	19.1	62.1	18.2
4	75.9	37.9	85.2	19.9
5	68.9	44.4	80.3	14.7
overall	**61.2**	**38.8**	**81.6**	**18.2**

with the glasses that outperforms high quality and well-controlled external references. We anticipate noise due to motion, temperature acclimatization, and external IR sources to become more pronounced under social, ambulatory conditions. These initial findings support the idea that large scale, naturalistic data collection are a fundamentally different challenge than typical highly-controlled, heavily-repeated laboratory experiments, and that Captivates is positioned well to extend our research into these challenging, novel scenarios.

6.2 Design

As we've outlined above, a major objective of the Captivates project is to create a pair of glasses that significantly improves on aesthetics and comfort, while maintaining all day robust performance. As much as possible, we wanted to build something that would not alter behavior or interpersonal interaction; a pair of glasses that had the highest likelihood of fading into the background over the day.

To test this hypothesis, we purchased several other pairs of smartglasses for comparison and ran some initial surveys. The glasses we used are shown in Figure 24; they include well known smartglasses: Google Glass, which includes a suite of functionality, including an outward facing camera; Bose AR, which simply plays audio; the Snapchat Spectacles, which can take a picture on demand; and the GoVision video recording glasses. We chose these glasses as reference points because they bound interesting corners of our design space– all include electronics, but the Bose AR hides it completely, designed for broad appeal by a large corporate team with access to market research. Google glass embraces a tech-futuristic aesthetic; our prototype could be (generously) construed this way as well. Spectacles are well-designed, but have outward facing sensors like Captivates.

To compare against these products, we included our initial prototype (same functionality, unoptimized for aesthetics) and our final design. Our survey population included 101 participants; 63% female, and of all participants, 50% wear glasses daily and overall, skewed slightly towards 26-35 year-olds (57%). In addition to evaluating each pair of glasses, we rated participants on the SCS-R scale for Public Self Consciousness and Social Anxiety, to see if there were trends mediated by these psychometrics. Distributions along these axes were unimodal and no clear differences in ratings appear for high vs low groups in either category.

The results, shown in Figure 25, reveal some interesting trends. In the survey, a neutral option was also available to participants. We opted to exclude neutral ratings in these charts to highlight trends in strong reaction to each pair, but each rating sums to a count of 101 with neutrals included.

First, we notice that all of these glasses fare pretty poorly on style– even the most positively reviewed pair here received only 20% favorable ratings (Captivates received 18%). This is not surprising after our conversations with glasses manufacturers; glasses are a style accessory, and it's quite difficult to make a single pair that is generally well-liked. Our initial prototype performed notably worse than the rest, both as ranked for personal style and in general appearance. The final Captivates design performs comparably to these other products.

The Captivates glasses actually do quite well compared to the others along self-assessed behavioral dimensions; they perform quite competitively with scores of comfort wearing them in public, private, and during conversation, as well as in assessment of discomfort to others (only outperformed by the Bose AR glasses). This is a truly significant shift from the rankings of the prototype (where an overwhelming 85% are uncomfortable with the idea of wearing them in public or while conversing), and clearly justifies the increased effort if our goal requires us to minimize changes in social behavior, self-consciousness, and stress or anxiety.

These behavioral ratings do not follow from the style ratings; they instead appear more closely linked to notions of social norm violation (do they look normal, and is it clear that they encapsulated electronics). Along these dimensions, the Captivates fares quite well, alongside the Bose AR and the GoVision frames.

A reasonable conclusion is that self-assessed comfort in social situations is dominated by the stigma surrounding social norms (video recording and electronics) compared with personal fashion or style match, as long as the style is somewhat conservative and doesn't grab attention. That is a positive result; it validates the notion that a generic pair of glasses doesn't alter self-assessed behavior as long as its style is within the norm (there is minimal advantage to trying to tailor the glasses to individual style for research). It also validates the notion that going from an obvious prototype that is outside the norm and has obvious electronics to something that looks 'normal' and integrated has a huge impact on self-assessed behavior.

There is still obvious room for improvement. The Bose AR are the best anchor point in this survey, as they include minimal electronics that are not visible, they have no electronics or unusual features on the face of the glasses, and they've had commercial success. A strong majority of people (58%) are happy to wear these glasses in public and in conversation without perceived social stigma.

Our glasses still have room to grow to compete with Bose AR. That said, we have integrated significantly more sensors throughout the glasses, including in difficult to ignore areas of the front face. If we take a look at open-ended comments surrounding the Captivates frames, most of the critiques are centered around four main themes: about 18% noted either bulkiness or the exposed IR diode in front as chief dislikes; the next most common complaints (roughly 7%) noted the surface finish or the lack of lenses.

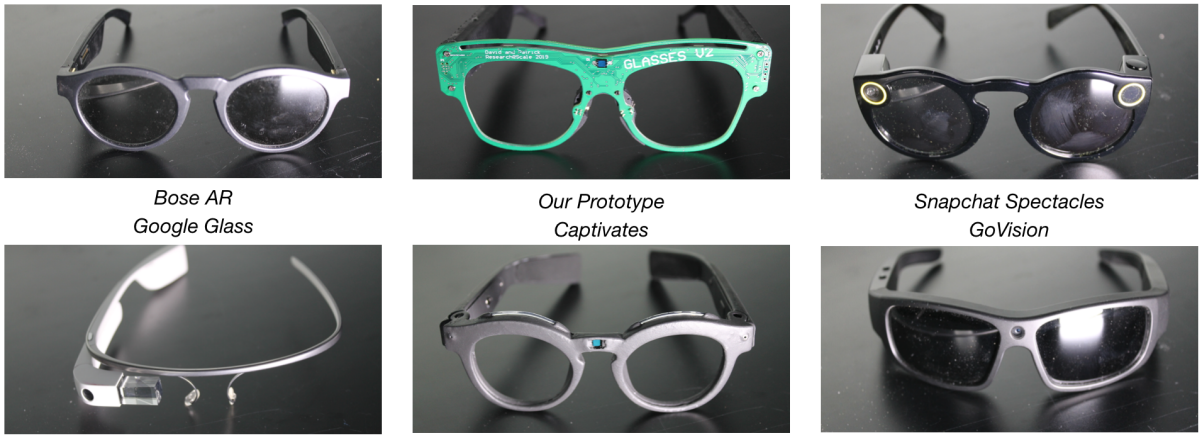

Fig. 24. Smartglasses Compared in Survey

Fig. 25. Glasses Survey Comparison Results

7 FUTURE WORK

The main issues identified in our validation survey were bulkiness, exposed IR diode, surface finish, and the lack of lenses. It's worth noting that the most common critique– bulkiness– is a major issue we dedicated a significant engineering effort to improve, and there is little left to optimize given the several sensors included in this design. With regards to the other issues, we will be addressing all of them– we are working with both lens suppliers and plastics suppliers, so we expect to resolve the lenses and improve surface finish issues; we have created an

'on-the-go' version of the glasses which does not have the exposed IR diodes or LED fiber optic accent when these are not required for a given study. With these changes in place, we believe the Captivates will fare even better than the confidence we can take away from our initial feedback, measuring across natural environments without altering user experience; that is a an achievement impossible with bulky prototypes. In our case, less than 5% of people were comfortable wearing our initial prototype in public, and it would drastically alter their behavior and their experience should they be forced to.

Even without these changes, our study indicates Captivates are socially acceptable and capable of large-scale naturalistic data collection. We are in the process of expanding our validation beta study efforts, and scaling production of our system for large scale studies. We're also working on secondary infrastructure– a supplementary, small, data logging device and other mobile friendly tools. Our next major step is to collect physiological responses across days, coupled with cognitive state self-assessment, from an at-home study. We'll use this study to (1) identify and improve any long-term usability issues, which we'll address before offering other researchers support in running studies with our platform, and (2) to collect the scale of data we need to inform our modelling work.

We will train probabilistic modelling algorithms to estimate user cognitive state from physiology once we've collected this data. Because of the irreducible uncertainty in mental state estimation, and the injective relationships that characterize psychophysiological monitoring, informed priors and probabilistic techniques will be necessary to build useful predictions of user experience; we are collaborating with developers on probabilistic programming languages to bring these techniques to bare on this data. These techniques have shown great promise compared to standard machine learning practices in the fields of computational cognitive science and causal modelling.

With models in place, we plan to use Captivates as a tool to study and inform design choices (for instance, how do phone notifications affect overall focus? How does a particular change in a user interface affect mental engagement?) and as a part of a personalized, adaptive ecosystem. For example, prior work has demonstrated the value of personalized learning policies [54]. With both real-time and long-term estimates of engagement, we can identify and adapt effective learning policies that are personalized to the learner.

8 CONCLUSION

With the Captivates project, we set out to build a system for long-term, in-the-wild psychophysiological monitoring at scale. We integrated many physiological sensors into a glasses form factor, particularly in challenging locations around the bridge of the nose. Our area of interest– naturalistic psychological measurement– dictates that our wearable be unobtrusive and socially acceptable. With that in mind, we focused heavily on an aesthetic that would minimize social anxiety and stress, as well as optimizing our comfort, weight, sensor selection, and battery life to let the user put them on without thought and forget about them all day long.

Our validation shows success in data quality and success in aesthetic design, in-so-much that people self-report that they would feel comfortable across public and private interactions, including conversation, while wearing them. It also validates that this was absolutely not true for our initial prototype, which 85% of people would not feel comfortable wearing in public, and demonstrates the value in our additional engineering effort to enable naturalistic study.

We believe to make meaningful progress on psychophysiological modeling, we not only need to move to naturalistic settings, but collect vast amounts of data. Once again, Captivates was designed to hit this mark– it is a scalable and robust solution, enabling large-scale, long-term studies with minimal administrative oversight. It is also open-source and available for other researchers to use and adapt.

Given these intentions, Captivates unique contributions are as follows:

- An open-sourced, manufacturable and scalable platform that is available for other researchers to use, audit, and modify. Captivates enable researchers to leverage a hardware platform for their specific use cases, and

can enable large-scale, trustworthy data sharing on physiology and mental experience without proprietary algorithms or APIs common to productized platforms.
- A product-like, comfortable, and aesthetically neutral design that can be used in real-world settings without altering naturalistic behavior. This is crucial for studying this physiology under as-realistic conditions as possible, and prototype-like designs alter self-reported behavior. These improvements require significant additional engineering effort.
- A robust form factor that is easy to use and lasts all day on a single charge. This is crucial for scaling trustworthy data collection without administrative overhead, and makes the glasses less intrusive into a participant's daily experience. Large scale data will be necessary to make any generalizations given the high levels of noise and uncertainty associated with real world settings.
- A platform to study understudied physiological indicators of face temperature, subtle head motion, and blinking in natural settings. These have been demonstrated in lab contexts, but naturalistic analysis of these indicators is novel. Furthermore, our comparison against standard camera-based techniques shows we significantly outperform them for blink tracking and spatially consistent temperature measurement, even in a relatively uniform working environment. Heart and respiration rate estimation from our head mounted IMU data should be possible as well. As a stand-alone device, or as part of a broader physiological measurement suite, Captivates serves as a power research tool for in-the-wild physiological and psychophysiological data collection and modeling.

In order to accomplish these goals, we spent a summer in China working with manufacturers and glasses designers, partnering with companies, and worked with sourcing agents. We pushed our design towards product and manufacturing in a way that is uncommon within academic research, but which we believe will become more prevalent in the near future. The ability to manufacture and quickly scale hardware design opens up opportunities for new research techniques. Captivates leverages this unique opportunity to scale data collection.

Captivates opens the door for large-scale data collection of physiological indicators, in a way that will enable better predictions of mental state in natural contexts. Our initial real-world pilot study yielded promising results and further data collection is planned to fully validate the platform. The immediate future will be spent collecting data and using that data to engineer and validate psychophysiological models that can estimate mental states accurately.

Success with these models would bring us a long way towards translating psychophysiological laboratory techniques into real-world insight. Captivates serve as an open-source bridge to that end. A strong model will enable Captivates to quantify the long-term psychological impact of our design decisions and provide real-time feedback for technologists interested in actuating a cognitively adaptive, user-aligned future.

The Captivates system is a key enabling piece of future responsive, adaptive environments. We envision an ecosystem of such technology, including dense arrays of distributed sensing and better actuation of shared spaces. With the Captivates project, we are one step closer in understanding the impact of our designs on human experience, and thus one step creating truly supportive, responsive environments.

ACKNOWLEDGMENTS

We would like to thank Adam Oranchak and Carl Allen from VSP Global on their support and mechanical design input. We would also like to thank Jie Qi and Bunnie Huang on organizing the MIT Media Lab Research at Scale Summer Program where we were given the opportunity to travel to Asia and meet and discuss our project with several designers and manufacturers. Finally, we would like to thank Gian C. Delfin and Emanuel Perez on their early effort in writing software to test a few of the system's functionalities.

REFERENCES

[1] 2020 (accessed November 3, 2020). *Base Station.* https://www.vive.com/us/accessory/base-station/
[2] 2020 (accessed November 3, 2020). *Focals by North.* https://www.bynorth.com/
[3] 2020 (accessed November 3, 2020). *FocusEDU.* https://www.brainco.tech/
[4] 2020 (accessed November 3, 2020). Meditation Made Easy. https://choosemuse.com/
[5] 2020 (accessed November 3, 2020). *Open Source Tools for Neuroscience.* https://openbci.com/
[6] 2020 (accessed November 3, 2020). *OpenThread.* https://openthread.io/
[7] 2020 (accessed November 3, 2020). *Quality RTOS Embedded Software.* https://www.freertos.org/
[8] 2020 (accessed November 3, 2020). *Thread.* https://www.threadgroup.org/
[9] 2020 (accessed November 3, 2020). *Your Everyday Smart Glasses.* https://vueglasses.com/
[10] 2021 (accessed April 2, 2021). *The world's first wearable eyewear that lets you see yourself.* https://jins-meme.com/en/
[11] Yomna Abdelrahman, Eduardo Velloso, Tilman Dingler, Albrecht Schmidt, and Frank Vetere. 2017. Cognitive Heat: Exploring the Usage of Thermal Imaging to Unobtrusively Estimate Cognitive Load. *IMWUT* 1 (2017), 33:1–33:20.
[12] Eugene Aidman, Carolyn Chadunow, Kayla Johnson, and John Reece. 2015. Real-time driver drowsiness feedback improves driver alertness and self-reported driving performance. *Accident Analysis & Prevention* 81 (2015), 8–13.
[13] Md Zahangir Alom, Tarek M Taha, Chris Yakopcic, Stefan Westberg, Paheding Sidike, Mst Shamima Nasrin, Mahmudul Hasan, Brian C Van Essen, Abdul AS Awwal, and Vijayan K Asari. 2019. A state-of-the-art survey on deep learning theory and architectures. *Electronics* 8, 3 (2019), 292.
[14] Amphenol 2018. *ZTP-135SR Thermometrics Thermopile IR Sensor.* Amphenol. Rev. H.
[15] Analog Devices 2020. *NDIR Thermopile-Based Gas Sensing Circuit.* Analog Devices.
[16] Celia Andreu-Sánchez, Miguel Ángel Martín-Pascual, Agnès Gruart, and José María Delgado-García. 2017. Looking at reality versus watching screens: Media professionalization effects on the spontaneous eyeblink rate. *PLoS One* 12, 5 (2017), e0176030.
[17] Pietro Aricò, Gianluca Borghini, Gianluca Di Flumeri, Alfredo Colosimo, Stefano Bonelli, Alessia Golfetti, Simone Pozzi, Jean-Paul Imbert, Géraud Granger, Raïlane Benhacene, et al. 2016. Adaptive automation triggered by EEG-based mental workload index: a passive brain-computer interface application in realistic air traffic control environment. *Frontiers in human neuroscience* 10 (2016), 539.
[18] Guha Balakrishnan, Fredo Durand, and John Guttag. 2013. Detecting pulse from head motions in video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 3430–3437.
[19] Tadas Baltrusaitis, Amirali Bagher Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. OpenFace 2.0: Facial Behavior Analysis Toolkit. *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)* (2018), 59–66.
[20] Juan A. Castro-García, Alberto J. Molina-Cantero, Manuel Merino-Monge, and Isabel M. Gómez-González. 2019. An Open-Source Hardware Acquisition Platform for Physiological Measurements. *IEEE Sensors Journal* 19 (2019), 11526–11534.
[21] Alper Cömert and Jari Hyttinen. 2015. Investigating the possible effect of electrode support structure on motion artifact in wearable bioelectric signal monitoring. *BioMedical Engineering OnLine* 14 (2015).
[22] Enrico Costanza, Samuel Inverso, Elan Pavlov, Rebecca Allen, and Pattie Maes. 2006. eye-q: eyeglass peripheral display for subtle intimate notifications.
[23] Rodney J. Croft and Roger J. Barry. 2000. Removal of ocular artifact from the EEG: a review. *Neurophysiologie Clinique/Clinical Neurophysiology* 30 (2000), 5–19.
[24] Erin Cvejic, Jeesun Kim, and Chris Davis. 2010. Prosody off the top of the head: Prosodic contrasts can be discriminated by head motion. *Speech Communication* 52, 6 (2010), 555–564.
[25] Ian Daly, Nicoletta Nicolaou, Slawomir Nasuto, and Kevin Warwick. 2013. Automated Artifact Removal From the Electroencephalogram: A Comparative Study. *Clinical EEG and neuroscience : official journal of the EEG and Clinical Neuroscience Society (ENCS)* 44 (05 2013). https://doi.org/10.1177/1550059413476485
[26] Linh C Dang, Gregory R Samanez-Larkin, Jaime J Castrellon, Scott F Perkins, Ronald L Cowan, Paul A Newhouse, and David H Zald. 2017. Spontaneous eye blink rate (EBR) is uncorrelated with dopamine D2 receptor availability and unmodulated by dopamine agonism in healthy adults. *Eneuro* 4, 5 (2017).
[27] Artem Dementyev and Christian Holz. 2017. DualBlink: A Wearable Device to Continuously Detect, Track, and Actuate Blinking For Alleviating Dry Eyes and Computer Vision Syndrome. *IMWUT* 1 (2017), 1:1–1:19.
[28] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition.* Ieee, 248–255.
[29] Elena Di Lascio. 2018. Emotion-Aware Systems for Promoting Human Well-being. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers.* 529–534.
[30] Adam Dunkels, Bjorn Gronvall, and Thiemo Voigt. 2004. Contiki-a lightweight and flexible operating system for tiny networked sensors. In *29th annual IEEE international conference on local computer networks.* IEEE, 455–462.
[31] Christoph Ellmer. 2017. OpenThread vs. Contiki IPv6: An Experimental Evaluation.

[32] Gianluca Di Flumeri, Pietro Aricò, Gianluca Borghini, Nicolina Sciaraffa, Antonello Di Florio, and Fabio Babiloni. 2019. The Dry Revolution: Evaluation of Three Different EEG Dry Electrode Types in Terms of Signal Spectral Features, Mental States Classification and Usability. *Sensors (Basel, Switzerland)* 19 (2019).
[33] Peter Gerjets, Carina Walter, Wolfgang Rosenstiel, Martin Bogdan, and Thorsten O. Zander. 2014. Cognitive state monitoring and the design of adaptive instruction in digital environments: lessons learned from cognitive workload assessment using a passive brain-computer interface approach. *Frontiers in Neuroscience* 8 (2014).
[34] Gabriele Gratton. 1998. Dealing with artifacts: The EOG contamination of the event-related brain potential. *Behavior Research Methods, Instruments, Computers* 30 (1998), 44–53.
[35] Kiran B. Hebbar, J. Dennis Fortenberry, Kristine Rogers, Robert Merritt, and Kirk Easley. 2005. Comparison of temporal artery thermometer to standard temperature measurements in pediatric intensive care unit patients. *Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies* 6 5 (2005), 557–61.
[36] Heimann Sensor 2013. *Application Note NDIR-Measurement.* Heimann Sensor.
[37] Javier Hernandez, Yin Li, James M Rehg, and Rosalind W Picard. 2014. Bioglass: Physiological parameter estimation using a head-mounted wearable device. In *2014 4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH)*. IEEE, 55–58.
[38] Javier Hernandez, Rob R Morris, and Rosalind W Picard. 2011. Call center stress recognition with person-specific models. In *International Conference on Affective Computing and Intelligent Interaction*. Springer, 125–134.
[39] David Hoppe, Stefan Helfmann, and Constantin A Rothkopf. 2018. Humans quickly learn to blink strategically in response to environmental task demands. *Proceedings of the National Academy of Sciences* 115, 9 (2018), 2246–2251.
[40] Stephanos Ioannou, Vittorio Gallese, and Arcangelo Merla. 2014. Thermal infrared imaging in psychophysiology: Potentialities and limits. In *Psychophysiology*.
[41] Shoya Ishimaru, Kai Kunze, Koichi Kise, Jens Weppner, Andreas Dengel, Paul Lukowicz, and Andreas Bulling. 2014. In the Blink of an Eye - Combining Head Motion and Eye Blink Frequency for Activity Recognition with Google Glass. *ACM International Conference Proceeding Series*, 150–153. https://doi.org/10.1145/2582051.2582066
[42] Shoya Ishimaru, Kai Kunze, Koichi Kise, Jens Weppner, Andreas Dengel, Paul Lukowicz, and Andreas Bulling. 2014. In the blink of an eye: combining head motion and eye blink frequency for activity recognition with Google Glass. In *AH '14*.
[43] Habib Karaki and Vasco Polyzoev. 2020 (accessed May 3, 2020). Demystifying Thermopile IR Temp Sensors. https://www.fierceelectronics.com/components/demystifying-thermopile-ir-temp-sensors-0
[44] Matthew A Killingsworth and Daniel T Gilbert. 2010. A wandering mind is an unhappy mind. *Science* 330, 6006 (2010), 932–932.
[45] Wolfgang Klimesch. 1999. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews* 29 (1999), 169–195.
[46] Nataliya Kosmyna, Caitlin Morris, Utkarsh Sarawgi, and Pattie Maes. 2019. AttentivU: A Biofeedback System for Real-time Monitoring and Improvement of Engagement. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (2019).
[47] Saroj Lal and Ashley Craig. 2005. Reproducibility of the spectral components of the electroencephalogram during driver fatigue. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology* 55 2 (2005), 137–43.
[48] Celine Latulipe, Erin A Carroll, and Danielle Lottridge. 2011. Love, hate, arousal and engagement: exploring audience responses to performing arts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1845–1854.
[49] Hayley Ledger. 2013. The effect cognitive load has on eye blinking.
[50] Robert Makepeace and Julien Epps. 2015. Automatic task analysis based on head movement. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (2015), 5167–5170.
[51] Tota Mizuno and Yuichiro Kume. 2015. Development of a Glasses-Like Wearable Device to Measure Nasal Skin Temperature. In *HCI*.
[52] Nargess Nourbakhsh, Yuhuai Wang, and Fang Chen. 2013. GSR and Blink Features for Cognitive Load Classification. In *INTERACT*.
[53] Rafal Paprocki and Artem Lenskiy. 2017. What does eye-blink rate variability dynamics tell us about cognitive performance? *Frontiers in human neuroscience* 11 (2017), 620.
[54] Hae Won Park, Ishaan Grover, Samuel Spaulding, Louis Gomez, and Cynthia Breazeal. 2019. A Model-Free Affective Reinforcement Learning Approach to Personalization of an Autonomous Social Robot Companion for Early Literacy Education. In *AAAI*.
[55] Francesco Pompei. [n.d.]. Temporal artery temperature detector.
[56] PyQT. 2012. PyQt Reference Guide. (2012). http://www.riverbankcomputing.com/static/Docs/PyQt4/html/index.html
[57] Mirko Raca, Lukasz Kidzinski, and Pierre Dillenbourg. 2015. Translating head motion into attention-towards processing of student's body-language. In *Proceedings of the 8th international conference on educational data mining*.
[58] Fabian Ramseyer and Wolfgang Tschacher. 2014. Nonverbal synchrony of head-and body-movement in psychotherapy: different signals have different associations with outcome. *Frontiers in psychology* 5 (2014), 979.
[59] Umair Rehman and Shi Cao. 2017. Augmented-Reality-Based Indoor Navigation: A Comparative Analysis of Handheld Devices Versus Google Glass. *IEEE Transactions on Human-Machine Systems* 47 (2017), 140–151.
[60] Harry T Reis and Samuel D Gosling. 2010. Social psychological methods outside the laboratory. (2010).

[61] Sergio Romero, Miguel Angel Mañanas, and Manel J. Barbanoj. 2008. A comparative study of automatic techniques for ocular artifact reduction in spontaneous EEG signals based on clinical target variables: A simulation case. *Computers in biology and medicine* 38 3 (2008), 348–60.

[62] Soha Rostaminia, A. Mayberry, D. Ganesan, Benjamin M Marlin, and Jeremy Gummeson. 2017. iLid: Low-power Sensing of Fatigue and Drowsiness Measures on a Computational Eyeglass. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1 2 (2017).

[63] Phattarapong Sawangjai, Supanida Hompoonsup, Pitshaporn Leelaarporn, Supavit Kongwudhikunakorn, and Theerawit Wilaiprasitporn. 2020. Consumer Grade EEG Measuring Sensors as Research Tools: A Review. *IEEE Sensors Journal* 20 (2020), 3996–4024.

[64] Zach Shelby, Klaus Hartke, and Carsten Bormann. 2014. The Constrained Application Protocol (CoAP). *RFC* 7252 (2014), 1–112.

[65] Alexander Shtuchkin. 2020. DIY Position Tracking using HTC Vive's Lighthouse. https://github.com/ashtuchkin/vive-diy-position-sensor.

[66] John A. Stern, Larry Walrath, and Rebecca Newberger Goldstein. 1984. The endogenous eyeblink. *Psychophysiology* 21 1 (1984), 22–33.

[67] Rainer Stiefelhagen. 2002. Tracking focus of attention in meetings. *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces* (2002), 273–280.

[68] Maja Stikic, Chris Berka, Daniel J. Levendowski, Roberto F. Rubio, Veasna Tan, Stephanie Korszen, Douglas Barba, and David Wurzer. 2014. Modeling temporal sequences of cognitive state changes based on a combination of EEG-engagement, EEG-workload, and heart rate metrics. *Frontiers in Neuroscience* 8 (2014).

[69] B. Tag, Andrew W. Vargo, Aman Gupta, G. Chernyshov, K. Kunze, and Tilman Dingler. 2019. Continuous Alertness Assessments: Using EOG Glasses to Unobtrusively Monitor Fatigue Levels In-The-Wild. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (2019).

[70] Giacomo Tanganelli, Carlo Vallati, and Enzo Mingozzi. 2015. CoAPthon: Easy development of CoAP-based IoT applications with Python. *2015 IEEE 2nd World Forum on Internet of Things (WF-IoT)* (2015), 63–68.

[71] Texas Instruments 2011. *TMP006/B Infrared Thermopile Sensor in Chip-Scale Package.* Texas Instruments. Rev. C.

[72] Chinchu Thomas and Dinesh Babu Jayagopi. 2017. Predicting student engagement in classrooms using facial behavioral cues. In *Proceedings of the 1st ACM SIGCHI international workshop on multimodal interaction for education*. 33–40.

[73] Garrick L. Wallstrom, Robert E. Kass, Anita Miller, Jeffrey F. Cohn, and Nathan A. Fox. 2004. Automatic correction of ocular artifacts in the EEG: a comparison of regression-based and component-based methods. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology* 53 2 (2004), 105–19.

[74] Chen Wang, Erik N Geelhoed, Phil P Stenton, and Pablo Cesar. 2014. Sensing a live audience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1909–1912.

[75] SV Wass, K de Barbaro, and K Clackson. 2015. Tonic and phasic co-variation of peripheral arousal indices in infants. *Biological Psychology* 111 (2015), 26–39.

[76] Nancy J Wei, Bryn Dougherty, Aundria Myers, and Sherif M Badawy. 2018. Using Google Glass in Surgical Settings: Systematic Review. *JMIR mHealth and uHealth* 6 (2018).

[77] Fadilla Zennifa, Junko Ide, Yukio Noguchi, and Keiji Iramina. 2015. Monitoring of cognitive state on mental retardation child using EEG, ECG and NIRS in four years study. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (2015), 6610–6613.