Beacon: Designing a Portable Device for Self-Administering a Measure of Critical Flicker Frequency

RAVI KARKAR, Computer Science & Engineering, DUB Group, University of Washington
RAFAL KOCIELNIK, Human Centered Design & Engineering, DUB Group, University of Washington
XIAOYI ZHANG, Computer Science & Engineering, DUB Group, University of Washington
JASMINE ZIA, Division of Gastroenterology, Medicine, University of Washington
GEORGE N. IOANNOU, Division of Gastroenterology, Medicine, University of Washington and Division of Gastroenterology, Veterans Affairs Puget Sound Health Care System
SEAN A. MUNSON, Human Centered Design & Engineering, DUB Group, University of Washington
JAMES FOGARTY, Computer Science & Engineering, DUB Group, University of Washington

Critical flicker frequency (CFF) is the minimum frequency at which a flickering light source appears fused to an observer. Measuring CFF can support early diagnosis of minimal hepatic encephalopathy (MHE), a condition affecting up to 80% of people with cirrhosis of the liver. However, adoption of CFF measurement in clinical practice has been hampered by the cost of a device for measuring CFF and the need for specialized training to administer the test. This paper presents *Beacon*, a portable, inexpensive device that enables people to measure their own critical flicker frequency. We adopt a mixed-methods approach to informing and evaluating the design of and potential opportunities for *Beacon*. We first report on a two-part formative study with 10 participants to evaluate the choice of certain parameters in the design of *Beacon*. We then report on a study of 41 healthy adults ranging from 18 to 99 years of age, finding that *Beacon* performs on par with Lafayette Flicker Fusion System, an established medical device, achieving a pearson correlation coefficient of 0.88. We finally report on a focus group with five hepatoligists who work with patients with cirrhosis of the liver, using our initial prototype development to examine their perspectives on potential opportunities and challenges in adoption of a device like *Beacon*. We discuss *Beacon* as an exploration of reframing critical flicker frequency measurement from a clinical screening tool into a self-administered self-tracking measure, thereby drawing upon and contributing to research in the health and personal informatics.

CCS Concepts: • **Human-centered computing → Empirical studies in ubiquitous and mobile computing**; *Ubiquitous and mobile computing systems and tools*; • **Applied computing → Health informatics**;

Additional Key Words and Phrases: Critical flicker frequency, hepatic encephalopathy, cirrhosis, self-tracking

ACM Reference Format:
Ravi Karkar, Rafal Kocielnik, Xiaoyi Zhang, Jasmine Zia, George N. Ioannou, Sean A. Munson, and James Fogarty. 2018. Beacon: Designing a Portable Device for Self-Administering a Measure of Critical Flicker Frequency. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3, Article 117 (September 2018), 27 pages. https://doi.org/10.1145/3264927

Authors' addresses: Ravi Karkar (rkarkar@cs.washington.edu), Xiaoyi Zhang (xiaoyi@cs.washington.edu), and James Fogarty (jfogarty@cs.washington.edu), Paul G. Allen School of Computer Science & Engineering, DUB Group, University of Washington, Seattle, WA 98195, USA; Rafal Kocielnik (rkoc@uw.edu) and Sean A. Munson (smunson@uw.edu), Human Centered Design & Engineering, DUB Group, University of Washington, Seattle, WA 98195, USA; Jasmine Zia (jzia@medicine.washington.edu), Division of Gastroenterology, Medicine, University of Washington, Seattle, WA 98195, USA; George N. Ioannou (george@medicine.washington.edu), Division of Gastroenterology, Medicine at University of Washington and Division of Gastroenterology and Health Services Research and Development at Veterans Affairs Puget Sound Health Care System, Seattle, WA 98195, USA.

2474-9567/2018/9-ART117 $15.00
https://doi.org/10.1145/3264927

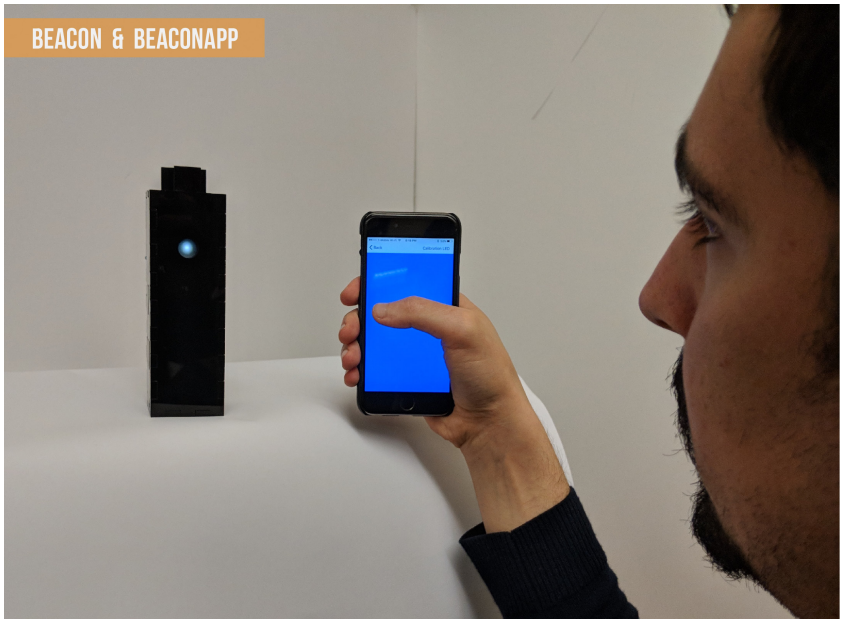


Fig. 1. *Beacon* — A portable, inexpensive, and self-administrable alternative for measuring critical flicker frequency (CFF) to enable timely detection of minimal hepatic encephalopathy (MHE). A person uses the *BeaconApp* to pair with the *Beacon* device over Bluetooth. The person can press anywhere on the blue screen to record their input (i.e., see a flickering light or see a fused light). In practice, the phone need not be facing the person and should be held outside of the person's view to avoid distraction.

1 INTRODUCTION

Chronic liver disease, including cirrhosis of the liver, is a major source of morbidity and mortality that is both preventable and currently underestimated [13]. In the United States alone, there are 3.9 million adults diagnosed with liver disease [2]. According to the Centers for Disease Control and Prevention, cirrhosis was the 12th-leading cause of death in the United States in 2013, accounting for more than 36,000 deaths [36].

Up to 80% of patients with cirrhosis will develop a spectrum of neurocognitive impairments known as *hepatic encephalopathy* (HE) [63]. HE fluctuates over time and occurs in a spectrum of severity ranging from *minimal hepatic encephalopathy* (MHE), which often goes undetected, up to *advanced/overt hepatic encephalopathy* (OHE), which can lead to coma and be life-threatening. MHE is associated with (1) subtle cognitive impairments and reduced quality of life [28]; (2) driving accidents [10] and falls [51]; (3) future development of OHE [8, 53] leading to hospitalization and even death [4]. If identified early, MHE can be treated and controlled relatively easily with medications (e.g., lactulose, rifaximin), avoiding these complications and reducing risk of progression [9, 50, 60].

Although practice guidelines recommend frequent screening of all cirrhotic patients for MHE [63], studies have demonstrated that cirrhotic patients are not being adequately screened. The need for expensive equipment and specialized training to operate that equipment are major barriers to broader adoption of early screening practices [7, 8]. Prior work has suggested public health efforts are needed to reduce the burden of cirrhosis, particularly among racial/ethnic minorities and among individuals at lower socioeconomic status [5, 55].

Critical flicker frequency (CFF) is a neurophysiological test that is an excellent candidate for timely screening for MHE. CFF is the minimum frequency at which a flickering light source appears fused to an observer.

Multiple studies have established that a healthy CFF of 40-45 Hz is reduced to <39 Hz in people with MHE [4, 35, 54, 57, 58, 62]. CFF has been shown to accurately detect MHE and, more importantly, to independently predict overall survival[4]. The accuracy of CFF in diagnosing MHE has been reported to be 80%, with a sensitivity and specificity of 65% and 91% [58]. However, measuring CFF currently requires specialized equipment, such as the *Lafayette Flicker Fusion System* (FFS, Lafayette Instrument Company, Lafayette, IN) [3]. To date, such specialized equipment has been used mostly as a research tool and is not available in routine clinical practice.

This paper presents *Beacon*, a portable, inexpensive, and self-administrable device for measuring critical flicker frequency. With *Beacon*, we aim to make CFF measurement more accessible to clinicians, patients, and caregivers. Although our goal is for *Beacon* to eventually be used as a clinical decision support tool, the initial research presented in this paper focuses on evaluating the performance of *Beacon* compared to an existing device among healthy participants. CFF measurement can be performed using different values for parameters like the intensity of the light source and the intensity of ambient light, and existing medical research has controlled for specific values of these parameters in prior studies. Because *Beacon* is motivated as a portable device, we cannot assume the same values used in prior medical research and need to examine how these values affect CFF measurement relative to a medical device like the Lafayette FFS. We therefore conducted two 10-participant formative data collection studies, examining the impact of the intensity of the light source and the intensity of ambient light on CFF measurement with *Beacon*. We found *Beacon*'s CFF measure was directly proportional to the intensity of the light source and inversely proportional to the intensity of the ambient light, which is in agreement with prior research. This formative data collection also allowed choosing appropriate values for these parameters in the *Beacon* device. We then conducted a 41-participant comparative study to examine CFF measurement with *Beacon* relative to CFF measurement with the Lafayette FFS, finding that *Beacon* measurements correlate to measurements by the gold-standard device with a Pearson's r of 0.88. Finally, we conducted a focus group with five senior hepatologists who work with patients with cirrhosis of the liver, using our initial prototype as a prompt to examine their perspectives on opportunities and challenges for adoption of a device like *Beacon*. Although none of the clinicians currently screen their patients with cirrhosis for MHE, they expressed desire to use a device like *Beacon* if it were available and easy to use.

The overall goal of this research is to design a portable device for self-administering CFF measurements. The specific contributions in this paper are:

(1) Developing *Beacon*: a portable, inexpensive, and self-administrable alternative for CFF measurement.
(2) Evaluating CFF measurement with *Beacon* against measurement with the gold-standard Lafayette FFS.
(3) Examining clinician perspectives on the current state of chronic care related to MHE, current practices regarding screening for MHE, and the potential opportunities and challenges for a device such as *Beacon*.
(4) Reframing CFF measurement from a clinical screening test to a self-administered self-tracking measure, thereby offering new potential for early identification and treatment of MHE before it progresses to the more severe forms.

2 BACKGROUND AND RELATED WORK

2.1 Hepatic Encephalopathy

2.1.1 Stages of Hepatic Encephalopathy. Cirrhosis of the liver is a major source of morbidity and mortality in the United States with 3.9 million adults diagnosed with liver diseases [2]. Hepatic Encephalopathy (HE) is a spectrum of neurocognitive impairments, and will occur at some point in up to 80% of patients with cirrhosis [63]. HE is one of the major unsolved complications of cirrhosis and there is a call to action from the medical community to increase diagnosis in its early stages [7]. One of the well-known classifications of the spectrum of HE is the West

Haven criteria, also known as the Conn score, based upon impairment in consciousness, intellectual function, and behavior [25]. It classifies HE into five grades or stages of severity, from Grade 0 (also known as minimal hepatic encdephalopathy (MHE)) to Grade IV (leading to coma and death), as seen in Figure 2.

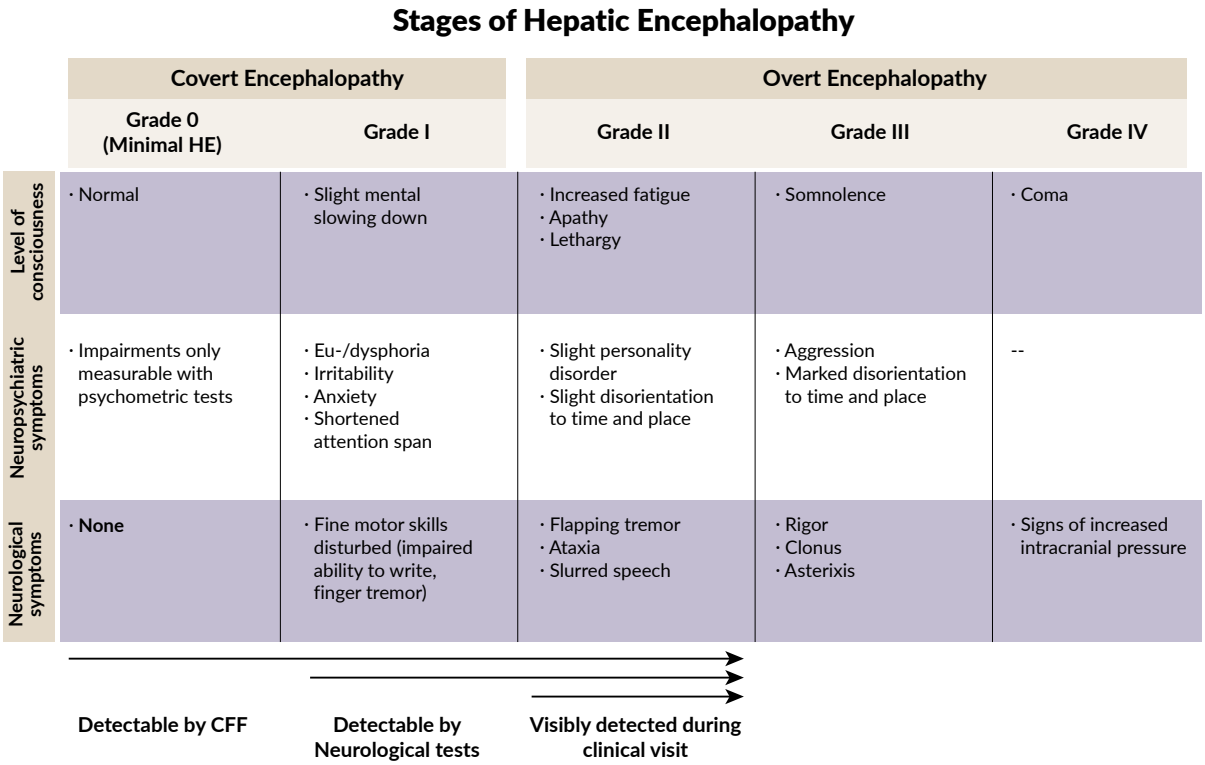


Fig. 2. The stages of hepatic encephalopathy. Current clinical screening practices only diagnose HE at Grade 1 or above. Our goal with *Beacon* is to identify Grade 0 HE (MHE) in individuals.

2.1.2 Minimal Hepatic Encephalopathy. MHE is a subtype of hepatic encephalopathy without manifest neurological symptoms, but with cognitive deficits [64]. The absence of neurological symptoms in its early stages and the difficulty of a patient with self-diagnosing cognitive impairments means that MHE remains undiagnosed in a vast majority of patients. MHE is associated with deficits in driving skills, reduction in quality of life, impairment of working capability, and predicting the subsequent onset of overt HE [45]. MHE is also a negative predictor for survival in patients with liver cirrhosis [64]. It is therefore important to detect MHE early when it is relatively straightforward to treat. In recent randomized, controlled studies, lactulose and rifaximin have improved the quality of life of patients with MHE and had a positive effect on their driving skills. Long-term therapy with medicine like lactulose or rifaximin significantly reduces the recurrence of episodic HE in patients who have previously had HE [64]. However, if not controlled, MHE can develop into overt HE with severe consequences, including leading to coma or death.

2.1.3 Screening for Minimal Hepatic Encephalopathy. Although practice guidelines recommend frequent screening of all cirrhotic patients for MHE [63], studies have demonstrated that cirrhotic patients are not being adequately

screened [7, 8]. This is in part because of the complicated and time-consuming nature of the two major categories of available screening tests: psychometric screening tests (i.e., tests measuring mental capabilities and behavioral style) and neurophysiologic screening tests (i.e., tests measuring function of the nervous system). In practice, this means that most patients with MHE are not diagnosed and going about their daily life unaware that they are more prone to driving accidents, falls, and other complications. By the time HE is identified in these patients, it has usually already progressed to Grade 1 or 2 (see Figure 2), with a person showing easily discernible signs such as asterixis (i.e., tremor of the hand when the wrist is extended) or disorientation in time and place. Vilstrup et. al. highlighted this lack of effective diagnosis of MHE stating, "*The diagnostic and treatment strategies for the overt form (Grade 2+) are generally well-outlined and agreed upon. However, it is the silent epidemic of covert HE (MHE) found in the majority of tested cirrhotic patients, which is an unmet need.*" [63]. With *Beacon*, we aim to help diagnose MHE at Grade 0 where the treatment is easier, patients are at lower risk of danger, and where early identification can significantly improve quality of life.

The Hepatic Encephalopathy Consensus Group at the World Congress of Gastroenterology in 1998 endorsed the psychometric hepatic encephalopathy score (PHES) as the gold standard for diagnosing Covert or Minimal HE [25] (Figure 3). Although the PHES is popular in Europe, it has failed to gain widespread acceptance in the United States, in part due to copyright issues and the lack of availability of testing material [49]. Several drawbacks of this testing system have also been identified over the past decade, including test administration and scoring in the outpatient setting [45]. Abnormalities in psychometric tests such as PHES can be caused by many other conditions that affect attention and concentration, such as sleep deprivation/insomnia, medications, alcohol, and drug use. In addition, performance in these tests is strongly affected by effort, training, age, interaction with the test administrator, literacy, numeracy, and education level [30, 59]. PHES and other psychometric tests are subject to learning effects, a key limitation in their suitability for self-monitoring of MHE. Because there are a limited number of variations of the test available, repeatedly administering the test on a regular basis is inappropriate and likely to undermine sensitivity of the test.

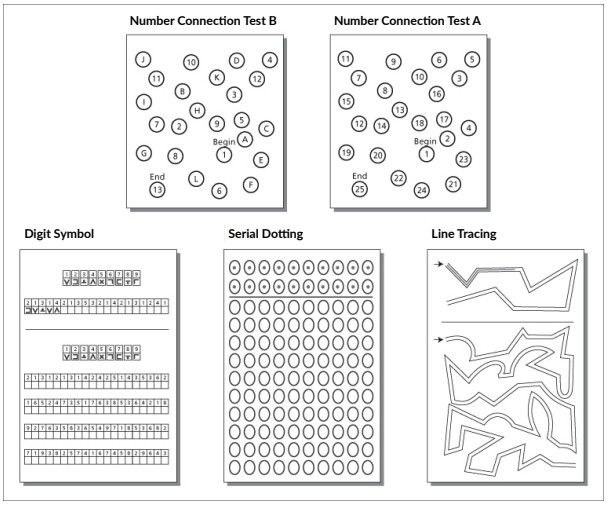
The above limitations make psychometric tests such as PHES less desirable for early diagnosis of MHE, leaving neurophysiological tests as the more reliable alternative. Critical Flicker Frequency (CFF) is one of the more promising alternatives, with a high specificity of 91% at diagnosing MHE [58]. However, the high cost of a device like the Lafayette FFS (about $3000) and need for special training to administer the measure have deterred its adoption.

2.2 Critical Flicker Frequency

Critical flicker frequency (CFF) is defined as the frequency at which an intermittent light stimulus (i.e., a blinking light) appears completely fused to the observer (i.e. steadily on, not flickering). It is a neuropsychological test of visual perception that has been used in detecting abnormalities in conditions ranging from visual signal processing (e.g., retinal gliopathy) to cognitive functioning. CFF is particularly apt for studying alterations in visual signal processing, and is suitable for the detection of arousal or attention abnormalities [52]. The CFF phenomena depends on both the brain and the retina, so transient effects in either can affect a person's critical flicker fusion frequency threshold. In addition to minimal hepatic encephalopathy, CFF has been applied to the study of neurological disorders including muliple sclerosis and Alzheimer's disease [52]. Important to our goals for self-monitoring for MHE, a neuropsychological test like CFF is also appropriate for repeated measurements (i.e., in contrast to the problematic learning effects expected with repeated administration of psychometric tests like PHES).

The CFF test has been in limited clinical and research use for over a decade, and its diagnostic accuracy has been evaluated in a number of reviews [6, 47, 48, 62]. Early research applying the CFF test established a set of standard testing conditions that have since been adopted by subsequent work [35, 62]. These conditions are also

Psychometric Hepatic Encephalopathy Score



Test	Description
Number connection test A (NCT-A)	Randomly dispersed numbers are to be connected with each other in serial order as quickly as possible.
Number connection test A (NCT-B)	Randomly dispersed numbers and letters are to be connected in alternating series (1-A-2-B...) as quickly as possible.
Digit-symbol	Digits from 1 to 9 are assigned respective symbols. Under each digit the corresponding symbol is to be written within a given time.
Serial dotting	Draw a dot inside each circle as quickly as possible.
Line tracing	A given line is to be traced as quickly as possible.

Fig. 3. Psychometric Hepatic Encephalopathy Score (PHES) consists of a battery of psychometric tests recommended for diagnosing minimal hepatic encephalopathy (example of the tests on the left, descriptions on the right).

implemented in commercially available devices, such as the Lafayette Flicker Fusion System[1] that we use as a gold standard in our research, the Hepatonorm Analyzer[2], and the Schuhfried Flicker Fusion Analyzer[3], each of which cost approximately $3000 USD. Standardized measurement of CFF is done in a quiet, semi-darkened room. A white light source is used to generate square waves with 50% duty cycle. Parameters of the test are chosen by the administrator, including the end points of the sweep (usually 25Hz and 55Hz) and step rate of sweep (usually 0.1, 0.5, 1, or 2Hz/sec). Starting at the lower end point (e.g., 25Hz), the light increases its frequency of flicker at a fixed step rate (e.g., 0.5Hz/sec). The person is asked to press a button once they perceive the light has stopped flickering. The light then starts from the highest end point (e.g., 55Hz) and decreases its frequency at a fixed step rate (e.g., 0.5Hz/sec). The person is asked to press the button once they perceive the light as flickering. This procedure is usually repeated 6 to 8 times to calculate the mean threshold value (CFF) and also estimate the standard deviation of the measurement. The whole procedure can therefore take about 10 to 15 minutes to complete [35].

In *Beacon*, we have aimed to preserve the core components of this procedure while adapting CFF to a form more appropriate for self-administration with a mobile device, with changes as detailed in the design section (3).

2.3 Personal Health Technologies

Health contributions in UbiComp research have taken a variety of forms, ranging from theoretical models to specific technology designs. One dimension of such research has considered personal informatics, exploring how people measure and track various aspects of their health and well-being with various goals in mind. Those findings are then distilled into models and frameworks that inform how tools can best support such practices. Models such as the Stage-Based Model [40] and the Lived Informatics Model [23] provide understanding of the

[1]Lafayette Flicker Fusion System (FFS)- http://www.lafayettelifesciences.com/product_detail.asp?itemid=36
[2]Hepatonorm Analyzer - http://www.nevolab.de/portfolio/hepatonorm-analyser/
[3]Schuhfried Flicker Fusion Analyzer - http://www.schuhfried.com/test/FLIM

tracking process barriers self-trackers are likely to encounter. Choe et al. further examined diagnostic self-tracking, explored the practices among people trying to answer questions about their health and the additional barriers they encounter due to lack of appropriate tools [14]. Within the health informatics community, Mamykina et al. contextualized self-tracking within chronic diseases and proposed the Sensemaking framework for chronic disease self-management[42]. Karkar et al. proposed a framework for self-experimentation in personalized health to guide people in answering questions about their health using self-tracking [32].

A second dimension of research has developed new approaches to better supporting existing medical practices. For example, systems have examined and supported tracking practices around specific needs in contexts that include mental health [11, 12], irritable bowel syndrome [31], diabetes [17, 24, 41], glucose monitoring[29], and sleep apnea [20]. A subset of such research has explored how to take specialized medical tests, currently confined to research or the clinic, into broader use. These include SymDetector for detecting sound based respiratory symptoms [61], contactless sleep apnea detection using smartphones [46], SpiroSmart and SpiroCall for enabling spirometry tests over a phone call [39], and PupilScreen for on-field assessment of traumatic brain injury [43].

A third dimension of research has explored how new approaches to capturing medical measures can lead to rethinking what the traditional measures mean and how the influx of self-tracked patient data can lead to new practices. For example, Kendall et al. explored daily blood pressure measurement and its recontextualization and reframing from the traditional clinical measure [34]. Researchers have also explored how traditional measures can be re-imagined in areas like food journaling [15, 16, 56] and weight measurement [33].

With *Beacon*, our contributions are in making the CFF measurement more accessible to patients and providers, initial exploration of the potential for new use of the measure in clinical practice, and discussing potential reframing of CFF from a clinical screening tool to a long-term self-tracking measure for chronic care.

3 DESIGN

3.1 Design Goal & Principles

3.1.1 Design Goal. With *Beacon*, we sought to create a portable, inexpensive, and self-administrable device for measuring the CFF threshold, thereby increasing access to the measure by clinicians and patients. We selected the Lafayette FFS, a medical gold standard device, as the reference device for us to model and compare performance against while building and testing *Beacon*. We adopted an iterative design process to arrive at the current version of the *Beacon* prototype.

3.1.2 Design Principles. Our design principles for *Beacon* were guided by current barriers to the adoption of the CFF measure in clinics. The key barriers we focused on addressing include: (1) limited access to testing devices due to cost constraints, and (2) the need for specialized training to use the device [8].

These led to our initial design principles that *Beacon* should be: (1) affordable enough enable adoption as a widely-used screening device, and (2) easy to use (i.e., requiring minimal training to operate). We also envisioned an additional use case for *Beacon*, enabling at-home CFF measurements for patient self-monitoring. This further emphasizes the above principle of minimal training, as people without medical training will be using the device. It also leads to the additional design principle that *Beacon* should have: (3) a portable form factor to allow easy at-home *self-administering*. These design principles are in addition to the core requirement that the *Beacon* prototype should perform as well as the existing medical devices that measure CFF.

3.2 Factors Affecting CFF Measurement

The CFF measure of an individual is not only affected by the neurological symptoms an individual might be experiencing, but also by the apparatus design and set up [37, 38]. We discuss some of the key factors that can affect an individual's CFF measurement, which therefore need to be addressed in a design, and how current medical devices like Lafayette FFS address these factors.

The key factors affecting an individual's CFF measure:

(1) The *intensity of the light source* used in the device is directly proportional to the measured CFF value (i.e., the brighter the light source, the higher the CFF value measured using that light).
(2) The *intensity of the ambient light* of the room has an inverse relationship with the measured CFF value (i.e., the brighter the room, the lower the CFF value measured in that room).
(3) The *viewing angle* also affects the CFF measure, with an ideal viewing angle being close to 0°. Peripheral vision has a greater sensitivity to movement, and hence a higher CFF value may be measured at other viewing angles.

Measuring CFF also requires choice of an appropriate *threshold detection algorithm*. A common algorithm used in CFF research is the *method of limits*, which is borrowed from psychophysical research and focuses on the influence and relationship between stimuli and the sensation and perception of these stimuli by an individual[26]. In the method of limits, a stimulus (i.e., light in the case of CFF) is presented and increased or decreased until it is perceivable by an individual. The primary parameter to be tuned in the method of limits is the *step rate* (i.e., the rate of change in the stimulus's ascent or descent). The design trade-off between step rate and effective CFF measurement is that of *reaction time* and *fatigue*. If the step rate is high (e.g., 2Hz/sec), even a small delay in reaction time would amount to a large error in the measured CFF. If the step rate is too low (e.g., 0.1Hz/sec), there will be an adverse impact of fatigue, which will likely compound during consecutive CFF measures.

3.3 Formative Design Work

We now discuss how these factors are addressed by current medical devices like the Lafayette FFS and by *Beacon*. Systems like the Lafayette FFS encase the light source in a viewing chamber (Figure 4A) and have the individual use a mask (Figure 4B) with the device. The viewing chamber allows even a low intensity light to be clearly visible in the dark chamber, and the mask ensures that ambient light is completely blocked out (Figure 4 right image), thereby controlling for the effect of both variables. The viewing chamber also has two circular cut-outs for the eyes, which are aligned with the light sources, ensuring that the viewing angle is close to 0°.

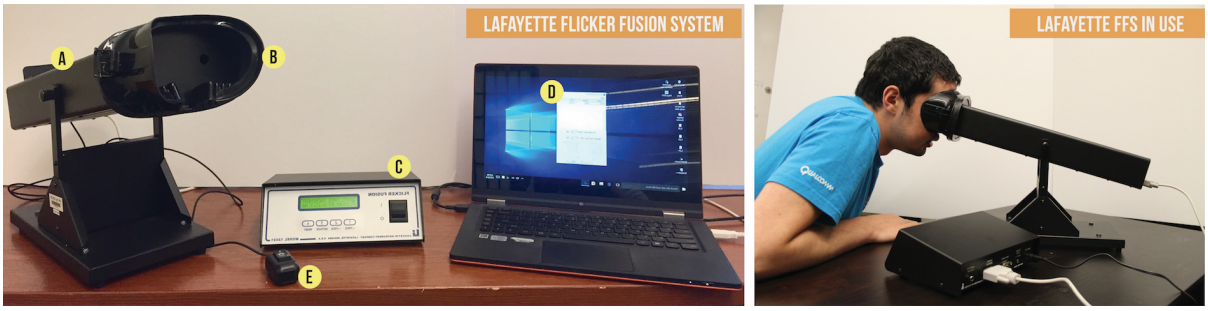


Fig. 4. Left: The Lafayette Flicker Fusion System (Lafayette FFS) has three components: a viewing chamber with the light stimulus (A, B), a controller (C), and a software program to record results (D). The clicker (E) is connected to the viewing chamber and records a person's input (i.e., indicating they see a flickering light or a fused light). Right: To use the Lafayette Flicker Fusion System, a person presses their face against the mask covering the viewing chamber so as not to allow any outside light. The person then focuses on the light inside the viewing chamber and uses the clicker to record their input.

We evaluated the feasibility of using existing capabilities of a modern phone to measure CFF and faced prohibitive limitations in choosing an appropriate light source, which deterred us from further pursuing that

direction. Although many phones include an LED used as a flash, no modern phones provide the API access necessary to generate frequency modulations to administer a CFF measurement. Additionally, standard phone's display does not have a high enough refresh rate to generate the necessary frequencies at 50% duty cycle.

The goal of our first prototype was to create a less expensive CFF device which can be operated using a phone (i.e., replacing the controller and computer set up with an application on the phone). Initially, by still using a viewing chamber design, we addressed the three factors in the same manner as Lafayette FFS. We began with a cardboard prototype to understand the set up we would need to build *Beacon*, such as what components are required and how the components might be housed (Figure 5 B). Upon finalizing the initial electronics (i.e., LED, microprocessor, power source), we upgraded to an acrylic box that allowed us to manipulate the distance between the light source and the eyes by sliding the LED inside the viewing chamber along pre-cut slots (Figure 5 C). It was during these design explorations that we realized that we could make the device portable by removing the viewing chamber. This led us to the first portable prototype of *Beacon*, which housed the LED and held the circuit inside it (Figure 5 D). However, the removal of viewing chamber would require examining the three factors described above to ensure the device still functions as intended. We conducted the formative studies described in section 4 to determine the value of the intensity of the light source and ambient light. We also added four calibration lights, red LEDs, as guiding lights to assist an individual in self-aligning themselves horizontally and vertically to be centered with the white light (Figure 5 E & 6).

The phone application, *BeaconApp*, connects with the RFduino over Bluetooth and allows an individual to self-administer the CFF measurement. It can be used to run the protocol to measure CFF (i.e., the method of limits), turn the calibration lights on or off, and view results. To start the measurement and record input, a person can press anywhere on the screen (i.e., a person does not look at the screen while measuring CFF).

Towards our design principles, we offer details about the affordability, portability, and ease-of-use of the current *Beacon* prototype. The total cost of the prototype, including the microprocessor, LED, diffuser, wiring, and acrylic for the case is less than $50 USD. Our current prototype's external dimensions are 140 x 40 x 40 mm. For a device like Lafayette FFS, the process of setting up the device includes plugging in cables across the viewing chamber, controller, and computer, starting the components and software, and then starting the measurement process. In contrast, the set up process for *Beacon* includes turning the device on and launching *BeaconApp* on the phone to begin taking a measure.

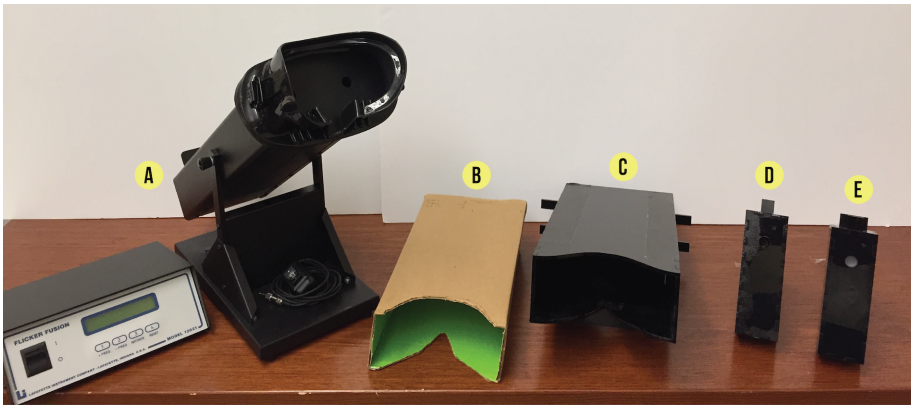


Fig. 5. Stages of the iterative design process for *Beacon*. Left: (A) Lafayette FFS, the golden standard device that informed our design of *Beacon*. Right: (B) Cardboard prototype, (C) Acrylic prototype with adjustable distance between light source and eyes, (D) portable prototype, and (E) portable prototype with guiding lights for horizontal and vertical alignment.

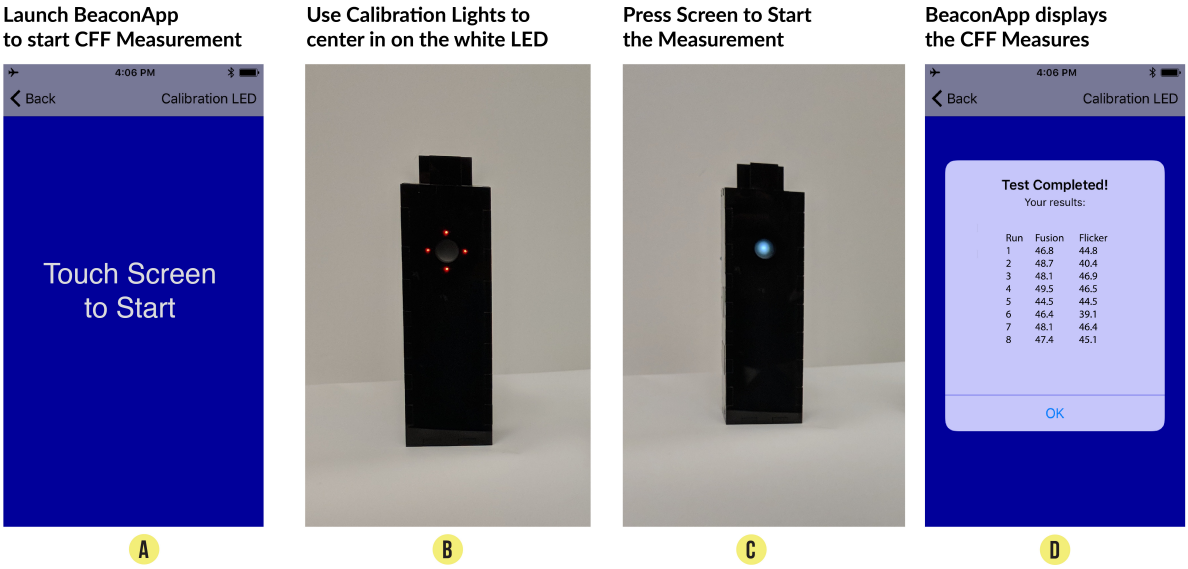


Fig. 6. (A) To begin the process of a CFF measurement, a person launches *BeaconApp*. They can press the 'Calibration LED' button on the top right of screen to initiate calibration step, or skip the step and start the CFF measurement by pressing anywhere in the blue region. (B) To ensure an appropriate viewing angle, *Beacon* uses two sets of two red LEDs forming a target around the central white LED. During the calibration phase, the red LEDs are visible through a tiny holes only when looking straight at them (i.e., at a viewing angle close to 0°). (C) The CFF measurement is conducted using the method of limits, with the LED starting the ascending step at 25.0Hz and increasing at a fixed rate of 0.5Hz/sec. A person presses on the anywhere on the screen when they see it as no longer flickering. The same process is repeated in the descending step, decreasing from an initial 55.0Hz. (D) After a person has repeated the ascending and descending steps a predefined number of times (8 times in image), they are presented with their results of each step. The CFF is calculated as a mean of all the steps.

3.4 Implementation Details

Beacon is built using off-the-shelf components and software with minor modifications. We use a RFduino microprocessor (RFD22102), selected because of it's small form factor and two-way communication API allowing for rapid prototyping of the mobile application to control the device. We selected a wide range white LED (C503D-WAN-CCBEB151 - luminous intensity from 28cd to 64cd) paired with a milky white diffuser in front of it to be the light source. The RFduino generates 50% duty cycle square waves (i.e., where pulse remains high for half of the period and low for the other half of the period) ranging from 25.0Hz to 55.0Hz at a step rate of 0.5Hz/sec (implemented as 0.1Hz/0.2sec) for the ascending and descending phase of CFF measurement.

Although past experiments have used a wider range of frequencies, most commonly between 20.0Hz and 60.0Hz, we selected 25.0 to 55.0Hz because (1) we expect all participants (including healthy individuals) to fall comfortably between those end points based on past medical research in MHE, and (2) a tighter range means a shorter run time which leads to less fatigue. Similarly, past experiments have used step rates ranging from 0.1Hz/sec to 3Hz/sec[22, 35]. We selected 0.5Hz/sec as a compromise in the trade-off between accuracy and fatigue.

3.5 Using *Beacon* & *BeaconApp*

We now briefly illustrate how *Beacon* and *BeaconApp* work through a scenario following a person who needs their CFF measure taken, either at the clinic or at home.

The person would sit in a dimly lit room in a chair next to a table. They place *Beacon* in front of them, on the table, at eye height. They then turn on *Beacon*, and open the associated *BeaconApp* on their phone. Parameters for the measurement are preset per clinician instructions (i.e., minimum and maximum frequencies of the sweep, step rate, number of repetitions). Administering a measurement then follows these steps:

- Turn on the red calibration lights to ensure the person's view is aligned with the center (Figure 6B).
- Press on the *BeaconApp* screen (Figure 6A), which automatically turns off the calibration lights and then begins the CFF measurement procedure.
- The central white LED starts at the lowest frequency of the sweep (i.e., perceived as flickering) and then increases at a fixed step rate until the person presses the application screen to indicate that they see it as fused (i.e., no longer flickering) (Figure 6C).
- Upon pressing the screen again, the central white LED starts at the highest frequency of the sweep (i.e., perceived as fused) and then decreases at the same step rate until the person presses the application screen to indicate that they see it as flickering.
- The ascending and descending measure are repeated several times per clinician instructions.

Voice prompts are provided by *BeaconApp* at each step of this procedure, guiding a person through the process without requiring they look at the phone application. Upon completing the final repetition, a person sees a screen which shows them their CFF measure from this session (Figure 6D).

4 STUDY DESIGN & RESULTS

A multi-disciplinary team of researchers across computer science, design, and medicine have contributed to the design and development of *Beacon*. Although our end goal with *Beacon* is for it to be used as a clinical decision support tool, we first need to evaluate whether its performance is comparable to existing CFF measurement devices. In this paper, we demonstrate that *Beacon* performs on par with an existing gold standard device (Lafayette FFS) using healthy study participants. In addition to comparing the two devices, we also conducted a focus group with hepatologists regarding their current practices for diagnosing HE and possible opportunities and challenges for *Beacon*. This focus group complemented the expertise of the medical researchers on the team. In particular, we were able to solicit additional expert perspectives on the opportunities afforded by *Beacon*'s portable and easy-to-use form factor, including the prospect of patients using *Beacon* for at-home self-monitoring of their HE.

We adopt a mixed-methods approach to informing and evaluating the design of and potential opportunities for *Beacon*. We conducted three studies, with each study informing the next:

(1) Two *formative data collection studies* examining the impact of the intensity of the light source and the intensity of ambient light on CFF measurement with our current *Beacon* prototype.
(2) A *comparative study* examining CFF measurement with our current *Beacon* prototype relative to CFF measurement with the Lafayette FFS.
(3) A *focus group* with hepatologists to understand current practices regarding screening and treatment of hepatic encephalopathy and their perspectives on the potential opportunities and challenges of integrating of a device like *Beacon* into their clinical care.

The formative and comparative study consisted of self-reported healthy individuals above 18 years of age. We recruited healthy participants because the goal of the studies was to compare the designs of the devices and not to perform a medical diagnosis. As noted in our introduction, the effectiveness of CFF as a diagnostic test for

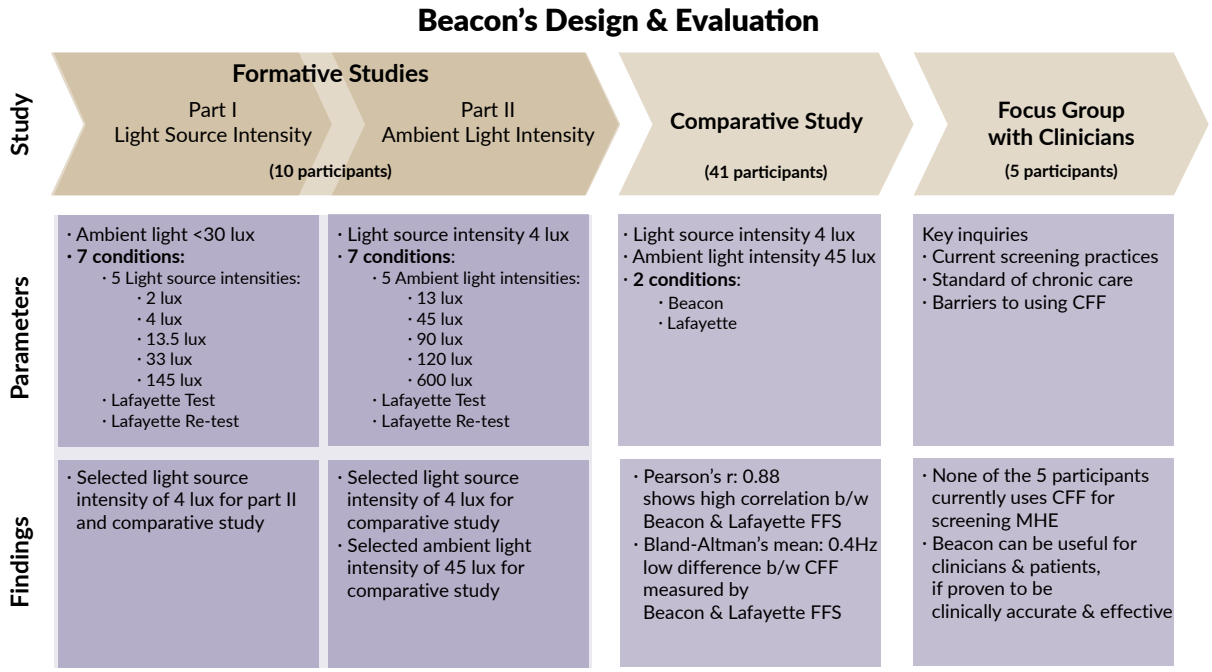


Fig. 7. Series of studies used to evaluate the design of and potential opportunities for *Beacon*. Formative studies assisted in selecting appropriate values of light source intensity and ambient light intensity for *Beacon*. Comparative study provided evidence of *Beacon*'s comparable performance to Lafayette FFS in measuring CFF of healthy individuals. Focus group provided information about current screening practices and potential impact of *Beacon* in screening and treating patients with HE.

MHE is already proven, with a reported accuracy of 80%, with a sensitivity and specificity of 65% and 91% [58] Our recruitment methods and study protocols were reviewed and approved by our University's Human Subjects Division under IRB number 00001222, *Evaluating a Novel Flicker Frequency Device*.

4.1 Formative Studies to Configure *Beacon*'s Parameters

Due to *Beacon*'s portable design, we cannot use the same values of intensity of the light source and intensity of ambient light that have been used in prior medical research (i.e., similar to the designs which consisted a viewing chamber). We therefore conducted two formative studies to examine the impact of these two parameter on the measurements of CFF by *Beacon*. For each intensity parameter, we selected five different values to examine their impact on the CFF measurement. Conducting a fully balanced study including two parameters with five levels each would be overly burdensome and fatigue-inducing for participants. To circumvent this problem, we examined each parameter independently and utilized the findings from the first to inform the design of the second. We initially measured the impact of light source intensity while controlling the ambient light intensity. We then used a fixed value of the light source intensity while varying ambient light intensity. We recruited the same participants for both parts of these formative studies.

4.1.1 Part 1: Impact of Light Source Intensity on Measured CFF.

Study Parameters. To examine how varying the intensity of the light source impacts the CFF measured by

Beacon, we collected data using **five levels of intensity — 2 lux, 4 lux, 13.5 lux, 33 lux, and 145 lux**. The intensity levels are not evenly spaced because (1) visual perception is not linear, and (2) we selected them subjectively to cover a wide range from dim to extremely bright. Approximately 15 different intensities were visually inspected in a dimly lit room, to emulate conditions we might expect in the wild, and from among them five we chose 5 for this formative study. Light intensities were measured using a light meter sensor (Amprobe LM-200LED) pressed against the diffuser of *Beacon* in a completely dark room. The different intensities were generated using the same white LED by varying the value of the resistor.

We configured both *Beacon* and the Lafayette Flicker Fusion System to use the previously-described method of limits algorithm, with a range between 25.0Hz and 55.0Hz, and a step rate fixed at 0.5Hz/sec.

Procedure. The study was conducted in an empty office room on the University campus. Participants first filled out a pre-task form consisting of their legal name, age, gender, and contact information. Participants were then introduced to the Lafayette FFS and *Beacon*. Because learnability of the system was not an aim of this study, a researcher guided each participant through measuring their CFF using both devices for at least two ascending and two descending measures until they were comfortable and confident with the measurement process. The study began after the participant reported they understood how the CFF measure worked and the devices functioned. The ambient light of the room was measured by putting the sensor next to *Beacon* and facing the wall. The ambient light was less than 30 lux across all 10 participants with a mean of 16.9 lux, median of 15.35 lux and standard deviation of 5.44 lux.

To mitigate any carryover effect (e.g., fatigue, learning), the study included **seven conditions** administered in a completely randomized order. Five of the conditions were the varying intensities of the *Beacon* light source, and the other two conditions were the Lafayette FFS (i.e., Lafayette Test and Lafayette Retest). We included the collection of two measures with the Lafayette FFS to observe any variance of measured CFF within the same participants. For each of the seven conditions, participants were asked to do six ascending and six descending runs of the method of limits, yielding **12 measures per condition**. Although CFF measurements can be conducted with six to eight runs, we decided to collect 12 to account for possible outliers in the measurements due to lapses in attention, delays in response time, or some other unforeseen reasons. Participants recorded their input using either a study phone with *BeaconApp* installed and paired with *Beacon* or the switch associated with the Lafayette FFS (Fig. 4 C). Participants were instructed to take as many breaks as they needed during the study. The study session lasted between 60-75 minutes.

Recruitment. We recruited 10 healthy participants through campus mailing lists. Of the 10, six reported themselves as male and four as female. Participants were between the ages of 18 and 36 years, with a mean of 27.1 years, a median of 25.5 years, and standard deviation of 4.9 years. Participants were provided with a $15 gift card as a compensation for participating in the study.

Data Cleaning. Collecting 12 measures per condition allowed us to discard extreme measures which may have been caused by reaction time, confusion, or some other factor. This use of a *trimmed* or *truncated mean* is common to obtain a more robust statistic. For each of the seven conditions we discarded the four most extreme CFF measurements (i.e., two lowest and two highest), then averaged the remaining values per participant. We combined the two Lafayette FFS measures to form the Lafayette_Avg condition to use as a baseline comparison value.

Results. The goal of this part of the formative study was to examine the impact of light source intensity on the measured CFF and select an appropriate value for *Part 2* of the formative study. We examine the mean CFF value per condition (i.e., across participants) to understand the impact of light source intensity. We found that

145 lux (highest intensity) had a mean CFF of 43.01Hz, and 2 lux (lowest intensity) had a mean of 36.96Hz. The trend in Figure 8, shows that measured CFF is directly proportional to the *Beacon*'s light source intensity and is in alignment with prior research and our understanding [22]. The difference between the CFF measured using 145 lux and 2 lux intensities is 6.05Hz, or 15.7% of the Lafayette FFS's average mean (38.49Hz).

We also observed that there was low variability within the conditions, indicating consistent measures. Examining the variation within the seven conditions, we observed low CFF variability between participants ($\sigma_{\bar{x}}$= std error = $\frac{\sigma}{\sqrt{n}}$: 0.58Hz to 0.91Hz). We observed slightly lower variability for Lafayette device, test, re-test, and average ($\sigma_{\bar{x}}$: 0.56Hz to 0.60Hz) as compared to *Beacon* with its five intensities ($\sigma_{\bar{x}}$: 0.74Hz to 0.91Hz).

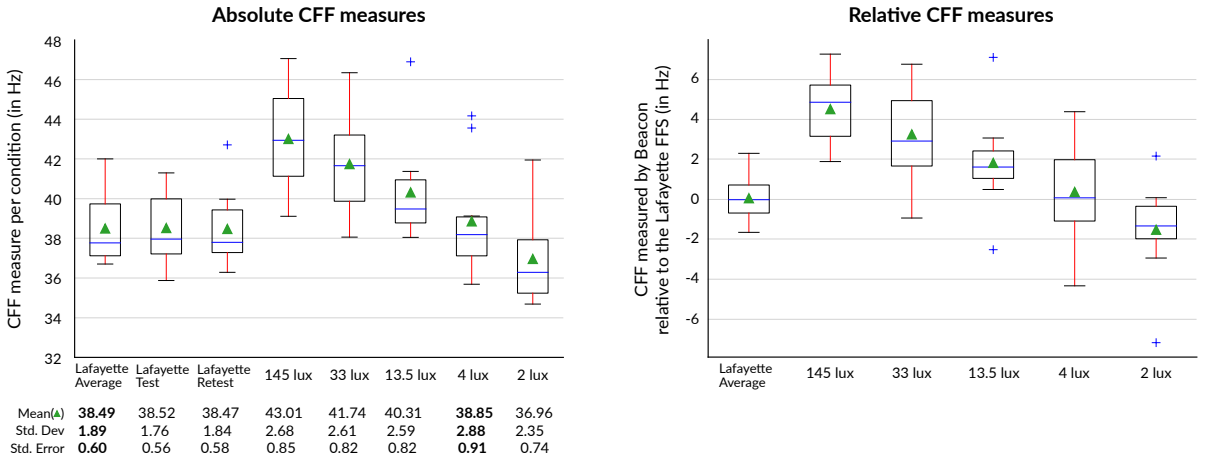


Fig. 8. Analysis of formative study *Part 1* examining the impact of light source intensity on measured CFF. The **blue** lines in the plot represent the median; the **green** triangles, the mean. Left: The absolute values of the 7 conditions (5 light source intensities, Lafayette test, and Lafayette retest) alongside a new calculated measure called the "Lafayette average" obtained by combining the test and retest scores. The plot shows a trend that CFF value is directly proportional to light source intensity. The table underneath the plot shows the corresponding descriptive statistics. Right: The values of light source intensities relative to the Lafayette average (using that average as the baseline). Light source intensity value of 4 lux, which is the closest to Lafayette average, was chosen for the remaining studies.

For *Part 2* of the formative study, we wanted to select a value of the light source intensity which would produce CFF measures closest to the Lafayette FFS. We observed the intensity of 4 lux had a mean CFF of 38.85Hz and standard error of 0.91Hz, which was closest to Lafayette_Avg mean of 38.49Hz and standard error of 0.60Hz. Based on these observations, we selected light source intensity of 4 lux as the study parameter for *Part 2* of the formative study.

4.1.2 Part 2: Impact of Ambient Light Intensity on Measured CFF.

Study Parameters. To examine how varying the intensity of ambient light impacts the CFF measured by *Beacon*, we collected data using **five levels of intensity — 13 lux, 45 lux, 90 lux, 120 lux, and 600 lux**. The different ambient light intensities were achieved by combining and manipulating the settings of multiple light bulbs capable of three different brightness settings (GE LED A21 3-Way Lamp). During the study, the participant sat on a chair at a table adjacent to a wall of the room, facing the wall, with all of the light sources behind the participant. We selected the ambient intensities based on how easy it would be to replicate the lighting conditions in a home

or a clinic. For perspective, 600 lux is similar to recommendations for office illumination while 45 lux is similar to recommended illumination for a relatively dark public area, such as a parking lot at night [1]. The light intensities were measured using a light meter sensor placed next to *Beacon* and facing the participant (i.e., facing the light sources). We used the same algorithm, range, and step rate as *Part 1* of the formative study. The light source intensity for *Beacon* was set at 4 lux as informed by the findings from *Part 1*.

Procedure. The study followed an identical procedure to that in *Part 1*, wherein participants were introduced to both the systems and completed practice runs before starting the study. The study consisted of seven conditions, using *Beacon* with the five varying intensities of ambient light of the room and using the Lafayette FFS device twice, with 12 measures per condition.

Recruitment. The 10 participants who participated in the previous study were recruited again for *Part 2*. Participants were provided with a $15 gift card as a compensation for participating in the study.

Data Cleaning. *Part 2* followed the same data cleaning process as *Part 1*.

Results. In this this part of the formative study, we examined the effect of ambient light intensity on the measured CFF and select an appropriate value for the comparative study. We observed that 600 lux (highest ambient light intensity) had the lowest mean CFF of 37.19Hz, while 13 lux (lowest ambient light intensity) had the highest mean CFF of 38.69 lux. The trend in Figure 9, shows that measured CFF is indirectly proportional to the ambient light intensity for *Beacon*, and is in alignment with prior research and our understanding [22]. The difference between the CFF measured using 600 lux and 13 lux intensities is 1.5Hz, or 3.85% of the Lafayette FFS's average mean (38.91Hz).

Examining the variation within the seven conditions, we observed low CFF variability between participants ($\sigma_{\bar{x}}$= std error = $\frac{\sigma}{\sqrt{n}}$: 0.58Hz to 0.73Hz). We observed similar variability for Lafayette FFS test, re-test, and average ($\sigma_{\bar{x}}$: 0.59Hz to 0.73Hz) as compared to *Beacon* with the five ambient intensities ($\sigma_{\bar{x}}$: 0.58Hz to 0.72Hz). Comparing differences between the highest and lowest intensities across the two parts of our formative studies, we observe that changing light source intensity has greater impact on resulting CFF measurement (a difference of 6.05Hz between the maximum and minimum, or 15.7% of mean) than changing ambient light intensity (a difference of 1.5Hz between the maximum and minimum, or 3.85% of mean).

For the comparative study (4.2), we desired a value of the light source intensity and ambient light intensity such that CFF measured by *Beacon* would be similar to the Lafayette FFS. Although ambient intensity of 13 lux was objectively the closest value to Lafayette_Avg (38.69Hz to 38.91Hz, difference of 0.22Hz), we decided to select **45 lux** for the comparative study because we deemed it was a much more reasonable level of ambient lighting for clinicians and patients to be able to achieve in their respective environments, and the impact on the CFF was small and not clinically meaningful (38.41Hz to 38.91Hz, a difference of 0.5Hz compared to a difference of 0.22Hz at 13 lux).

4.2 Comparative Study to Evaluate *Beacon*

The formative studies guided our decision of selecting appropriate light source and ambient light intensity parameters for *Beacon* to calibrate the measured CFF with the existing Lafayette FFS medical device. Using the selected parameters, we next conducted a comparative study to examine the CFF measured by *Beacon* relative to Lafayette FFS. Because our primary goal is to measure device performance, building upon prior research validating the potential for CFF measurement in diagnosis [58], we focus this comparison on healthy volunteer participants. Future research can then build upon this device validation to explore clinical usage with appropriate

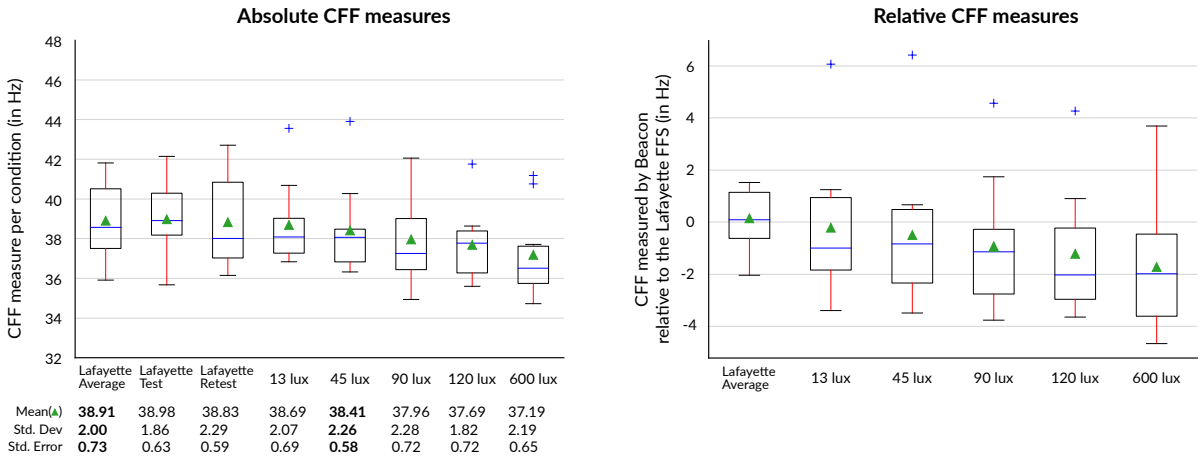


Fig. 9. Analysis of formative study *Part 2* examining the impact of ambient light intensity on measured CFF. The **blue** lines in the plot represent the median; the **green** triangles, the mean. Left: This plot shows the absolute values of the 7 conditions (5 light source intensities, Lafayette test, and retest) alongside a new measure called Lafayette average which combines the test and retest score. The plot shows a trend that CFF value is indirectly proportional to ambient light intensity. Table underneath the plot shows the corresponding descriptive statistics. Right: This plot shows the values of ambient light intensities relative to Lafayette average by using it as the baseline. Ambient light intensity of 45 lux was chosen for the comparative study since it was deemed to be a more easily achievable ambient light setting in clinics and homes.

patients. Similarly, we do not collect psychometric tests such as PHES because we do not expect them to be informative with healthy volunteer participants (i.e., there is no diagnosis to be made). Future clinical trials of *Beacon* could collect additional such measures as appropriate.

4.2.1 Study Design. We used the same algorithm, range, and step rate as the formative studies. Based on the previous studies, we set the light source intensity at 4 lux and ambient light intensity at 45 lux. We conducted a within subjects study with two conditions, using *Beacon* and using Lafayette FFS, counterbalanced across participants. For each of the **two conditions**, participants were asked to do eight ascending and eight descending runs of the method of limits, yielding **16 measures per condition**.

4.2.2 Procedure. To control the ambient light, the study was conducted in empty rooms with no windows. Similar to the formative studies, participant information was collected and they were introduced to both the systems and given a couple practice runs before starting the study.

4.2.3 Recruitment. 43 healthy participants, 18 years or older, were recruited for the comparative study using mailing lists, fliers, and sign up sheet at a retirement community. 24 participants reported themselves as female and 19 as male. Participants were between the ages of 18 and 99 years with a mean of 41.7 years, median of 28 years, and standard deviation of 25.3 years.

4.2.4 Data Cleaning. Data cleaning for the comparative study involved: (1) removing two participants due to missing data, (2) using an adaptive algorithm to discard outliers, and (3) testing the normality of collected data to justify use of specific statistical test.

We removed data for two participants (P33 and P35) from the analysis because we could not obtain their CFF measures using *Beacon*, leaving us with 41 participants for analysis. P33 (87 y.o.) was unable to see flicker using *Beacon* at any frequency for unknown reasons, their CFF measure using the Lafayette FFS was 31Hz. P35 (88 y.o.) who self-reported as being legally blind in one eye, had a very low CFF measure of 13Hz (using Lafayette FFS) which we were unable to capture without updating the *BeaconApp* during the session.

We collected 16 measures for each of the two conditions (i.e., *Beacon* and Lafayette FFS). Because of the potential for erroneous measurements (e.g., due to distractions), prior work using CFF has implemented a constraint where measures per participant should be collected repeatedly until the standard deviation is less than 3Hz [22]. We followed an adaptive algorithm for discarding high and low extremes based on a similar constraint:

- If max-sd > 3Hz, discard two extreme measures — lowest and highest.
- If max-sd still > 3Hz, repeat previous step.
- Terminate if number of measures is < 8 per condition.

We tested the normality of the data using D'Agostino and Pearson's test that combines skew and kurtosis to produce an omnibus test of normality [18, 19]. The normality test has the null hypothesis the sample is not normal, which means a $p \gg 0.05$ implies that the sample is not significantly different from a normal distribution (i.e., sample is normal). The normality test for the lafayette CFF measures (coef=1.529, p=0.465) and *Beacon* CFF measures (coef=3.033, p=0.219) suggested that normality was a reasonable assumption.

4.2.5 Results. The goal of the comparative study was to measure the performance of *Beacon* as compared to Lafayette FFS. As discussed in the introduction, the effectiveness of CFF for diagnosing MHE is already proven [58]. Demonstrating that *Beacon* performs on par with Lafayette FFS is a first step toward our goal of providing an accessible clinical decision support tool for measuring CFF.

We first examine the performance of our adaptive algorithm and then present our analysis on the cleaned data. For us to consider the adaptive algorithm effective, it should (1) be less sensitive to outliers than original data, and (2) still give a reasonable central tendency. As we can see from Table 1, before discarding any extremes the maximum standard deviation for Lafayette FFS is 5.51Hz and for *Beacon* is 8.59Hz (i.e., before applying our algorithm). After discarding 6 extremes, we still have 10 measures per condition and find the maximum standard deviation is much more similar, at 2.78Hz for Lafayette and 2.93Hz for *Beacon*. During this process the mean CFF values of Lafayette FFS changed from 37.60Hz to 37.58Hz and for *Beacon* they changed from 38.31Hz to 38.01Hz. Both of these difference are negligible, showing that the algorithm is removing outliers without affecting the mean. We also noted that while using the adaptive algorithm on Lafayette FFS, only 1 participant needed two extremes discarded, compared to *Beacon* where 6 participants had two extremes discarded, 5 had four discarded, and 1 had six discarded extremes. Although this did not impact the means, it suggests there is more internal variation in the measures taken by *Beacon* when compared to Lafayette FFS.

We took the mean CFF measured per participant in the two conditions and used correlation analysis to evaluate the consistency in measured CFF values between the devices (Fig. 10). Despite the normality test results suggesting reasonable assumption of normality, we present results from both Pearson and Spearman correlation analyses for greater transparency (i.e., because Pearson assumes normality while Spearman is non-parametric). We observed a Pearson correlation coefficient of **0.88** ($p \ll 0.001$) and Spearman correlation coefficient of **0.84** ($p \ll 0.001$), both of which correspond to strong correlations.

We used a Bland-Altman plot to better understand the agreement and the expected limits of difference between the CFF measurements taken by *Beacon* and Lafayette FFS. Also know as a difference plot, Bland-Altman plot is ideal for comparing two measurement techniques (or devices) that each produce some error in their measures and is extensively used to evaluate the agreement among two different instruments or measurement techniques [21, 22, 27]. Figure 10 shows the Bland-Altman plot on the right. The X-axis represents the mean of the CFF measurements for both devices and the Y-axis represents the absolute difference between the measurements

Table 1. Descriptive statistics for the Adaptive Algorithm used in the Comparative study data cleaning for each device (in Hz). The goal of the algorithm is to reduce maximum standard deviation while having minimal impact on the mean CFF. As seen here, using the adaptive algorithm *Beacon* achieved a maximum standard deviation of 2.93Hz compared to 2.78Hz achieved by Lafayette FFS. The mean CFF remains unaffected with a difference of only 0.29Hz in *Beacon* when not using and using the algorithm and 0.02Hz in Lafayette FFS.

	Lafayette FFS			***Beacon***		
Without discarding extremes	Mean CFF: 37.6033			Mean CFF: 38.31033		
	Max SD	Min SD	Mean SD	Max SD	Min SD	Mean SD
	5.5104	0.6743	1.6992	**8.4901**	0.7314	2.6949
After discarding 2 extremes	Mean CFF: 37.5870			Mean CFF: 38.1118		
	Max SD	Min SD	Mean SD	Max SD	Min SD	Mean SD
	2.7150	0.5943	1.2974	**7.5569**	0.5967	1.8686
After discarding 4 extremes	Mean CFF: 37.5787			Mean CFF: 37.9726		
	Max SD	Min SD	Mean SD	Max SD	Min SD	Mean SD
	2.5332	0.5006	1.0690	**6.0237**	0.4097	1.3807
Adaptive until SD per condition < 3Hz	Mean CFF: 37.5787			Mean CFF: 38.0145		
	Max SD	Min SD	Mean SD	Max SD	Min SD	Mean SD
	2.7837	0.6529	1.5801	**2.932**	0.7082	1.8197

taken by the two devices. The plot includes the line for the mean difference between the measurements (0.40Hz between Lafayette and *Beacon*) and the 2 lines showing the 2s (1.96 standard deviation) limits of differences between the measurements (also called 95% limits of agreement) which span from -3.27Hz to +4.07Hz for *Beacon* and lafayette. *The limits of agreement tell us that the difference in CFF measured by Beacon and Lafayette will be at most ±3.67Hz for 95% of the measurements.* A recent comparison of different CFF protocols using the same device reported limits of agreement from ±3.02Hz to ±6.74Hz [22]. Given that prior comparison of results from the same device using different algorithms, our results are promising when comparing two devices.

Our participant sample spanned a wide age distribution (18-99), which allowed us to examine the possible impact of age on the difference in CFF measurements by *Beacon* and Lafayette devices. Although we counterbalanced the order of measurement using each device, we also wanted to check if ordering had any systematic impact on differences in measurements. To examine both, we performed a regression analysis. We regressed the age and measurement order on the difference in CFF measurement between both devices. We found no significant impact of age ($\beta = -0.0043, SE = 0.013, p = 0.746$) or order ($\beta = -0.0171, SE = 0.619, p = 0.978$).

4.3 Focus Group with Clinicians to Understand Current Practices and Opportunities Surrounding Screening for MHE

The *Beacon* research team includes a gastroenterologist and a hepatologist, who provided medical expertise to support our development of *Beacon*. Our initial prototype and results demonstrated technical feasability of *Beacon*. To complement this, we gathered clinical perspectives on the acceptability and utility of such a device. Consultations with additional practitioners helped us better understand current diagnostic practices for HE, barriers to adoption of CFF, provider interest in a device like *Beacon*, and barriers they might envision to successful adoption. As the new form factor might allow new use cases for CFF, we were eager to explore provider views on more routine screening for MHE, such as patient use of *Beacon* for at-home self-monitoring of HE. Because HE

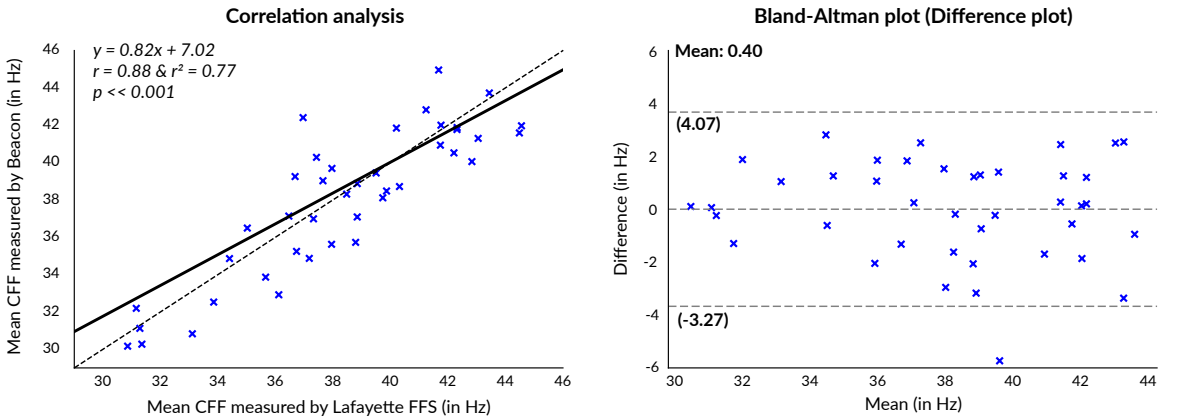


Fig. 10. Analysis of comparative study evaluating the performance of *Beacon* when compared to the Lafayette FFS. Left: Regression analysis shows a strong correlation between the CFF measure by *Beacon* and Lafaytte FFS with a Pearson's R of 0.88. Right: The Bland-Altman plot shows the mean difference between *Beacon* and Lafayette FFS to be 0.4Hz with a maximum difference of at most ±3.67Hz for 95% of the measurements.

primarily occurs among patients with end-stage liver diseases, treatment is often managed by specialists such as hepatologists. Consequently, we recruited five senior hepatologists at a medical school to participate in a focus group about challenges and opportunities associated with the adoption of a tool like *Beacon*.

4.3.1 Procedure. The focus group was conducted at a medical school for the convenience of the clinicians. Before starting the group discussion we collected background information about the clinicians, including title and years of experience treating patients with cirrhosis and HE. The focus group was structured around three key areas of inquiry in order to understand the current state of:

- Screening practices for patients with HE (e.g., At what stage are patients screened? How are they screened? What is the clinician's satisfaction with the screening process?)
- Standard of chronic care for patients with HE (e.g., What is the advice given? What does routine care include? Do patients do any self-tracking of their condition?)
- Barriers to using CFF as a screening option for MHE (e.g., Why they do or do not use CFF? What benefits would they anticipate if CFF were an easily available measure? Why might they decide not to use CFF?)

Participants also completed a short paper-based questionnaire during the focus group to collect individual responses to specific questions.

We offered $30 compensation to the clinicians as a token of appreciation, but each clinician declined the compensation and indicated they would prefer it spent on further research and development.

4.3.2 Recruitment. The five hepatologists were recruited through the medical school mailing lists. All the clinicians had more than 10 years of experience of seeing patients with cirrhosis and hepatic encephalopathy.

4.3.3 Data Analysis. The focus group discussion's audio was recorded and later transcribed by the researchers along with the questionnaire responses. The transcript was used to collect and organize the responses and reactions to the specific themes outlined in the three key areas of inquiry mentioned above. For any result based on the questionnaire, we report how many of the five participants were in agreement with the particular response.

4.3.4 Results. The goal of the focus group was to better understand current practices surrounding diagnosing HE and exploring possible avenues of integrating *Beacon* in the screening and chronic care process. We organize our findings from the focus group by the three key areas of inquiry outlined above. Quotes by specific clinicians are referred by C*X*.

Current Screening Practices & Diagnostic Workflow. All the participants reported that, as specialists (i.e., hepatologists), most of the patients with HE that they treat have been pre-diagnosed with potential HE and referred to them by either the primary care physician or a gastroenterologist. Each clinician reported using clinical evaluation (which can only diagnose grade 2 and above HE (Fig. 2)) as the primary method to screen patients for HE.

Clinicians agreed that many cases of MHE are missed or go undiagnosed until the condition has progressed to grade 1 or 2, where it is easily clinically diagnosable. While discussing MHE cases that go undiagnosed, clinicians mentioned time constraints during clinical visit as the primary reason for not using screening tests for MHE. However, clinicians also reported they would consider using a screening test for MHE if it were easy and quick enough to be administered by a medical assistant before a patient sees them, as highlighted by C4's comment *"if there was a tool that was accurate like the Stroop test that was available in the clinic which could be administered by a medical associate when they check their BP, heart rate, weight, test for encephalopathy."*

Describing topics covered during a routine clinical visit for the patient with HE, the clinicians reported conducting a clinical evaluation, going over their lactulose (medication) use, compliance with medication, and checking if they have any financial issues with the medication.

Current Standard of Chronic Care. Clinicians reported that the first advice they give patients diagnosed with HE was to stop driving or operating heavy equipment. C4 explained *"it is pretty hard for [patients] to assess themselves that they have enough impairment that [them driving with HE] is probably worse than walking outside the bar and thinking it is safe to drive"*. As a part of their treatment of the chronic condition, all clinicians reported prescribing patients Lactulose to restrict further deterioration and control HE. Lactulose is a colonic acidifier that works by decreasing the amount of ammonia in the blood. Although this does not cure HE, it does help improve the mental status of the patient.

Discussing commonly prescribed at-home chronic-care, clinicians ask patients to self-titrate lactulose to achieve a goal of about 2-3 *smooth bowel movements per day*.

Explaining how they track whether a patient's MHE is worsening or stable, C5 reported using some measure of cognitive ability like the number connection test (Fig. 3): *"sometimes I give a stack of number connection test and then use that to track patients who are subclinical [MHE], or if people like to play tablet or phone games which require some cognitive activity like I have a patient who plays Sudoku or something and they notice that their score is going down then that is a sign."* Clinicians also reported asking caregivers, family or professional, to update them in case of any noticeable change in personality or behavior.

Barriers to using CFF as screening tool. None of the five clinicians currently use CFF measurement as a part of their screening process. However, everyone expressed their desire to use *Beacon*, and thereby CFF, *if* it were proven to be *useful* (i.e., clinically capable of screening for MHE) and could be administered quickly. C1 explained *"there are lots of people who don't even have encephalopathy clinically [diagnosed] yet, so we could use it [Beacon]."* When asked about potential uses of *Beacon* beyond a clinician's screening of patients with MHE, C2 responded that *Beacon* could be a useful tool that provides some objectivity in what usually is a subjective evaluation: *"We spend a lot of time with family members who are trying to (convince) the patient that they shouldn't be driving and the patient doesn't believe it. But, if you have something objective it can be used to help them realize the situation."*

Although clinicians were excited about the potential to use *Beacon* in the clinic, they were more skeptical of at-home use by patients. Discussing potential use of *Beacon* at home by the patients, the clinicians raised liability concerns. C1 said *"What if we get the data and they got in a car accident? Do we have medical or legal liability since we had data suggesting that they had abnormal flicker [CFF]. I don't want that data."* Clinicians also expressed concerns with data overload burdening their already strained workload, such as patients coming to their clinical visit with another piece of information that the clinician needs to interpret. When discussing about the value of *Beacon* for caregivers looking after patients with HE, C2 said *"I have sense that with the population at hand their families will pretty much appreciate the objective data so that they can save the patient / loved ones [by preventing them from driving]."*

5 DISCUSSION

With *Beacon*, we seek to provide an accessible clinical decision support tool capable of measuring CFF. As the first step towards that goal, we conducted multiple studies to examine its feasibility for measuring CFF by comparing against an existing medical device. Although this is not the same as proving its diagnostic capabilities, it is a necessary precursor before testing the device with patients. We then collected impressions from experts to understand barriers to *Beacon*'s acceptance and provide feedback for future iterations. Our results indicate that *Beacon* has potential for use in screening and managing MHE, and that several barriers must still be overcome before it can be used as a screening tool in the clinic or at home. In addition to screening, development of a mobile, self-administered tool such as *Beacon* also enables new possibilities, such as observing CFF measurement trends over time to raise (i.e., observing a drop in CFF measurements over time) or reducing patient anxiety and concern (i.e., by observing stable or increasing CFF measurements over time).

5.1 Challenges & Opportunities in Designing *Beacon*

We designed *Beacon* as an effort to take a clinical measure which has been studied in research contexts and transition it in to clinical practice. Our comparative study demonstrated a strong correlation between the CFF measurements taken by *Beacon* and the gold standard Lafayette FFS. The reliability of *Beacon* makes new screening and monitoring possibilities available, such as providing patients with a device to measure CFF at home.

Through our studies, we also began refining the design in ways that will better support these opportunities. In our testing with participants at the retirement community, we discovered a design shortcoming in the *BeaconApp* wherein a person with limited dexterity or hand tremors might have unintended inputs due to accidentally pressing the screen in rapid succession. This resulted in two participants (P31 and P38) who had three and four measures respectively which were close to the range end-points of 25Hz and 55Hz (out of 16 measurements). We addressed this issue by introducing a 2-second delay between presses during which the screen remains inactive. This prevented any further misreports by other participants. Although adjustments like above can be easily implemented in a study, for a device like *Beacon* to be useful in practice it should be robust to adapt to a range of conditions. Although we observed that ambient light intensity has low impact on the measured CFF, one way to increase robustness for in-the-wild use would be to install components which automatically adjust the light source intensity by measuring the ambient light intensity (e.g., with a photocell).

During the comparative study, although the mean CFF reported by *Beacon* was close to that of the Lafayette FFS, we noticed slightly higher variance in the *Beacon* measures. The adaptive algorithm successfully removed the outliers and achieved the target of a <3Hz maximum standard deviation, affecting measures for 12 participants compared to just 1 for Lafayette FFS. We hypothesize that removal of the viewing chamber may have introduced this additional variance. We might explore increasing light source intensity and/or decreasing ambient light parameters to see if that reduces the variance.

Although greater measurement variability can be interpreted as an unreliable measure, we believe this can be offset by the richness of data which can be possibly collected by patients from their home. Daily or weekly CFF measures collected by a patient could provide an overview of their HE that can be useful to bring in during a clinical visit, instead of being limited to a single data point collected on the day of the visit. Our work on *Beacon* suggests the design question of how to meaningfully present the data collected by patient in a succinct and actionable manner.

The focus group provided valuable insights regarding the challenges and opportunities that might be expected in clinical adoption of *Beacon*. Clinicians emphasized that for *Beacon* to help them screen MHE in a clinical context, it would need to be accurate as well as easy and quick to administer. We also unpacked some of the potential opportunities and challenges associated with using *Beacon* as an at-home self-tracking tool for HE patients. Although *Beacon* may provide a crucial objective measure where none exists now, there remain important ethical and legal questions surrounding the collected data and its implications for medical practice that will need to continue to be examined. Because our participant pool was recruited from a single location and does not represent the diverse practices or needs of clinics around the country and the world, future work to move *Beacon* or a similar device into clinical use should engage with clinicians, staff, administrators, and patients in a variety of contexts.

5.2 Potential Limitations of a Single-Point Threshold Based Assessment

Medical studies have put forward the notion that a CFF of 39Hz is the threshold for MHE (i.e., if a person has CFF below 39Hz they likely have MHE) [35]. Based on our work, we highlight problems with the notion of a single-point threshold based assessment that does not take other factors into consideration.

As we discovered during our focus group, medical professionals are overburdened and do not have the bandwidth to perform long screening tests during a clinical visit. Even if the CFF measurements are conducted in clinics, medical professionals will not take more than the minimum number of observations to get a CFF measurement. If during those observations, the patient is distracted, or misunderstands the process, their CFF measurement will be flawed. To base their treatment on such a flawed measure would be erroneous.

Contrary to the notion that a healthy individual's CFF score should be >39Hz [35], the comparative results showed otherwise [44]. CFF scores between 30Hz and 35Hz were common among our participants at the retirement community, across both the devices. Part of this can be attributed to age-related degradation of sight reported by some participants (e.g., macular degeneration and astigmatism). Although the majority of the at-risk population might not be in that age group (65+ years old), it is crucial to understand that such thresholds or single-point assessments should be used with caution as like many other medical conditions, hepatic encephalopathy does not occur in a vacuum. Similarly, there can be other age or health related factors which can significantly reduce an otherwise healthy individual's CFF score.

We believe that by reducing the CFF measurement to a single-point assessment we are losing valuable insights like daily or weekly changes in condition, which can lead to better, more targeted health care interventions. *Beacon* offers a new potential approach to this limitation.

5.3 Monitoring CFF Trends as an Alternate to Single-Point Assessment

Prior to the development of *Beacon*, it was not practically feasible to collect CFF over time, and as such a single-point assessment was the only approach considered. To address the issues outlined in the above section, we propose reframing CFF from a clinical screening tool to a self-tracking measure collected at home.

As a preliminary investigation of the variability in CFF measurement over time, one member of the research team self-administered CFF measure using *Beacon* twice daily for one month: a morning measurement within 30 minutes of waking up and an evening measurement less than an hour before heading to bed. The measures were

found to be relatively stable as seen in Figure 11. To measure the sensitivity of *Beacon*, after the first two weeks, the morning measurement was switched to be later (i.e., after breakfast). We found that *Beacon* was able to detect that small shift in wakefulness as seen after the 14-day mark in Figure 11.

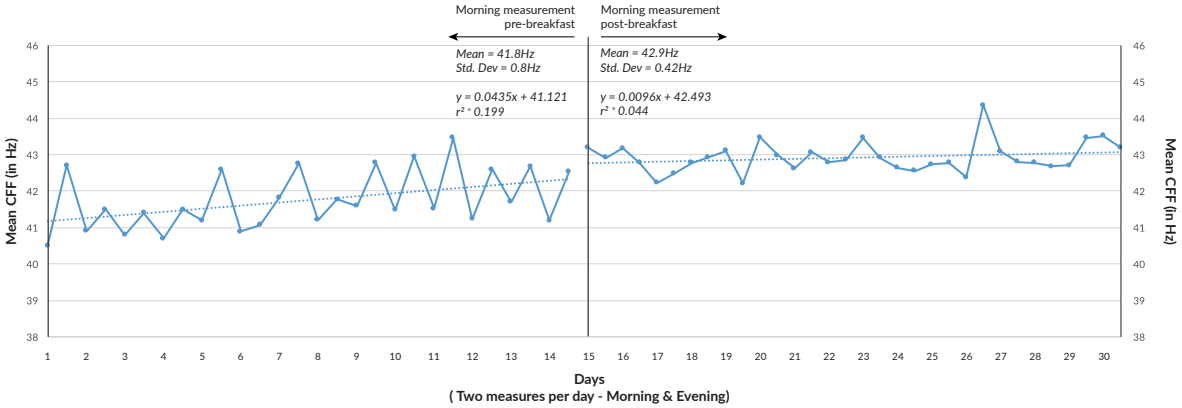


Fig. 11. Month-long CFF measurement data collected by member of the research team using *Beacon*. Measures were taken twice daily—morning and evening. After 2 weeks, the morning measurement time was pushed back to see if the change in wakefulness would be picked up by *Beacon*. *Beacon* successfully picked up the slight variation in wakefulness, providing evidence of potential for long term CFF tracking.

We envision a person with cirrhosis collecting their daily CFF measurements in relatively similar conditions (e.g., at a consistent point in their daily routine). Instead of comparing their data to a threshold value, we envision examining variation in a person's measurements over time. If a person's average CFF over a week drops by 1.5Hz or 2Hz, then they might reach out to their clinician or increase their dosage of lactulose. This shows an alternative possibility of thinking about the CFF measure. The frequency of CFF measurement in such a scenario might be agreed upon by the clinician and the patient. This practice can motivate further research in understanding the stability of CFF measurement for making treatment recommendations. Future work, enabled by a device like *Beacon*, can also explore what would be an ideal frequency of CFF measurements for people with different stages of HE, thus providing guidance to clinicians for tailoring treatment plans to a patient.

This new form of data and its representation raises interesting opportunities. It also raises important questions for patient-provider collaboration around such data in a patient's chronic self-management and about the ethics and legal implications of such data. Although we found support from the clinicians regarding using CFF in clinics and encouraging patients and caregivers to use CFF at home, there were still concerns regarding liability. This motivates further research investigating the longitudinal data collection of CFF measurements by patients, how to best communicate and present the measurements, and how to best enable collaboration around the data with clinicians.

Although our current focus with *Beacon* is to enable timely diagnosis of HE, CFF also has potential applications in other medical conditions like muliple sclerosis and Alzheimer's disease [52]. By making CFF measurement accessible, *Beacon* has the potential for impact across a range of medical conditions.

6 CONCLUSION

We designed and examined *Beacon*, a portable, in-expensive, and self-administrable device capable of measuring critical flicker frequency. We performed two formative studies with 10 participants each to validate the portable

design of *Beacon*, and we then performed a comparative study with 41 participants and found its performance to be comparable to an existing medical device. In a focus group with five hepatologists we found that *Beacon* would be valuable not just for clinicians but also for the patients and their caregivers if it was proven to be clinically effective. *Beacon* suggests new opportunities in cirrhosis and other conditions and motivates future work examining field deployment of such new forms of self-administrable medical diagnoses.

7 ACKNOWLEDGMENTS

We acknowledge Melody Kadenko, Eric Whitmire, Edward Wang, Alex Mariakakis, Daniel Epstein, and Kara Walter, and our study participants for their assistance in the research. This research has been funded in part by the University of Washington Innovation Research Award and the National Science Foundation under awards IIS-1553167 and IIS-1813675.

REFERENCES

[1] 2015. Recommended Light Levels (Illuminance) for Outdoor and Indoor Venues. (2015). https://www.noao.edu/education/QLTkit/ACTIVITY_Documents/Safety/LightLevels_outdoor+indoor.pdf

[2] 2015. Summary Health Statistics: National Health Interview Survey 2015. (2015). https://www.cdc.gov/nchs/fastats/liver-disease.htm

[3] 2018. Flicker Fusion System. (2018). http://lafayettelifesciences.com/products/flicker-fusion

[4] Javier Ampuero, Macarena Simón, Carmina Montoliú, Rodrigo Jover, Miguel Ángel Serra, Juan Córdoba, and Manuel Romero-Gómez. 2015. Minimal Hepatic Encephalopathy and Critical Flicker Frequency Are Associated with Survival of Patients with Cirrhosis. *Gastroenterology* 149, 6 (2015), 1483–1489. https://doi.org/10/f7x9d8

[5] Sumeet K. Asrani, Joseph J. Larson, Barbara Yawn, Terry M. Therneau, and W. Ray Kim. 2013. Underestimation of Liver-related Mortality in the United States. *Gastroenterology* 145, 2 (2013), 375–382. https://doi.org/10/f2kqcw

[6] Jasmohan S Bajaj. 2010. Review Article: The Modern Management of Hepatic Encephalopathy. *Alimentary Pharmacology and Therapeutics* 31, 5 (2010), 537–547. https://doi.org/10/cfxn47

[7] Jasmohan S. Bajaj. 2015. Diagnosing Minimal Hepatic Encephalopathy: From the Ivory Tower to the Real World. *Gastroenterology* 149, 6 (2015), 1330–1333. https://doi.org/10/cjc7

[8] Jasmohan Singh Bajaj, Ashkan Etemadian, Muhammad Hafeezullah, and Kia Saeian. 2007. Testing for Minimal Hepatic Encephalopathy in the United States: An AASLD Survey. *Hepatology* 45, 3 (2007), 833–834. https://doi.org/10/bsmscz

[9] Jasmohan S Bajaj, Douglas M Heuman, James B Wade, Douglas P Gibson, Kia Saeian, Jacob A Wegelin, Muhammad Hafeezullah, Debulon E Bell, Richard K Sterling, R Todd Stravitz, Michael Fuchs, Velimir Luketic, and Arun J Sanyal. 2011. Rifaximin Improves Driving Simulator Performance in a Randomized Trial of Patients With Minimal Hepatic Encephalopathy. *Gastroenterology* 140, 2 (2011), 478–487. https://doi.org/10/cdp62r arXiv:NIHMS150003

[10] Jasmohan S. Bajaj, Kia Saeian, Christine M. Schubert, Muhammad Hafeezullah, Jose Franco, Rajiv R. Varma, Douglas P. Gibson, Raymond G. Hoffmann, R. Todd Stravitz, Douglas M. Heuman, Richard K. Sterling, Mitchell Shiffman, Allyne Topaz, Sherry Boyett, Debulon Bell, and Arun J. Sanyal. 2009. Minimal Hepatic Encephalopathy Is Associated With Motor Vehicle Crashes: The Reality Beyond the Driving Test. *Hepatology* 50, 4 (2009), 1175–1183. https://doi.org/10/cw8k8c

[11] Jakob E Bardram, Mads Frost, Károly Szántó, Maria Faurholt-Jepsen, Maj Vinberg, and Lars Vedel Kessing. 2013. Designing Mobile Health Technology for Bipolar Disorder: A Field Trial of the Monarca System. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2627–2636. https://doi.org/10/cpqq

[12] Dror Ben-Zeev, Rachel Brian, Rui Wang, Weichen Wang, Andrew T Campbell, Min S H Aung, Michael Merrill, Vincent W S Tseng, Tanzeem Choudhury, Marta Hauser, John M Kane, and Emily A Scherer. 2017. Crosscheck: Integrating Self-report, Behavioral Sensing, and Smartphone Use to Identify Digital Indicators of Psychotic Relapse. *Psychiatric Rehabilitation Journal* 40, 3 (2017), 266–275. https://doi.org/10/ckh7

[13] Martin Blachier, Henri Leleu, Markus Peck-Radosavljevic, Dominique-Charles Valla, and Françoise Roudot-Thoraval. 2013. The Burden of Liver Disease in Europe: A Review of Available Epidemiological Data. *Journal of Hepatology* 58, 3 (2013), 593–608. https://doi.org/10/f2g4zr

[14] Eun Kyoung Choe, Nicole B. Lee, Bongshin Lee, Wanda Pratt, and Julie A. Kientz. 2014. Understanding Quantified-Selfers' Practices in Collecting and Exploring Personal Data. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2014)*. New York, New York, USA, 1143–1152. https://doi.org/bbpd

[15] Chia-Fang Chung, Jonathan Cook, Elizabeth Bales, Jasmine Zia, and Sean A Munson. 2015. More Than Telemonitoring: Health Provider Use and Nonuse of Life-Log Data in Irritable Bowel Syndrome and Weight Management. *Journal of Medical internet Research* 17, 8 (2015), E203. http://doi.org/10.2196/jmir.4364

[16] Chia-Fang Chung, Kristin Dew, Allison Cole, Jasmine K Zia, James Fogarty, Julie A Kientz, and Sean A Munson. 2016. Boundary Negotiating Artifacts in Personal Informatics: Patient-Provider Collaboration with Patient-Generated Data. In *Proceedings of the ACM Conference on Computer-Supported Cooperative Work and Social Computing*. https://doi.org/10/bn9c

[17] C Clar, Katharine D Barnard, Ewen Cummins, Pamela Royle, and Norman Waugh. 2010. Self-Monitoring of Blood Glucose in Type 2 Diabetes: Systematic Review. *Health Technology Assessment* 14, 12 (2010), 1–140. https://doi.org/10/cpqp

[18] Ralph D'Agostino and Egon S Pearson. 1973. Tests for departure from normality. Empirical results for the distributions of b_2 and $\sqrt{b}_1$. *Biometrika* 60, 3 (1973), 613–622.

[19] Ralph B d'Agostino. 1971. An omnibus test of normality for moderate and large size samples. *Biometrika* 58, 2 (1971), 341–348.

[20] Nediyana Daskalova, Danaë Metaxa-Kakavouli, Adrienne Tran, Nicole Nugent, Julie Boergers, John McGeary, and Jeff Huang. 2016. SleepCoacher: A Personalized Automated Self-Experimentation System for Sleep Recommendations. In *Proceedings of the Symposium on User Interface Software and Technology*. ACM, 347–358. http://doi.org/bwt4

[21] Katy Dewitte, Colette Fierens, Dietmar Stöckl, and Linda M Thienpont. 2002. Application of the Bland–Altman plot for interpretation of method-comparison studies: a critical investigation of its practice. *Clinical chemistry* 48, 5 (2002), 799–801.

[22] Auria Eisen-Enosh, Nairouz Farah, Zvia Burgansky-Eliash, Uri Polat, and Yossi Mandel. 2017. Evaluation of Critical Flicker-Fusion Frequency Measurement Methods for the Investigation of Visual Temporal Resolution. *Scientific Reports* 7, 1 (2017), 2–10. https://doi.org/10/gcmxfq

[23] Daniel A. Epstein, An Ping, James Fogarty, and Sean A. Munson. 2015. A Lived Informatics Model of Personal Informatics. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp 2015)*. 731–742. https://doi.org/bdsr

[24] Aojg Farmer, Oj Gibson, L Tarassenko, and A Neil. 2005. A Systematic Review of Telemedicine Interventions To Support Blood Glucose Self-Monitoring in Diabetes. *Diabetic Medicine* 22, 10 (2005), 1372–1378. https://doi.org/10/c75gjk

[25] Peter Ferenci, Alan Lockwood, Kevin Mullen, Ralph Tarter, Karin Weissenborn, and Andres T Blei. 2002. Hepatic Encephalopathy - Definition, Nomenclature, Diagnosis, and Quantification: Final Report of the Working Party at the 11th World Congresses of Gastroenterology, Vienna, 1998. *Hepatology* 35, 3 (2002), 716–721. https://doi.org/10/bmwzs9

[26] George A Gescheider. 2013. *Psychophysics: the fundamentals*. Psychology Press.

[27] Davide Giavarina. 2015. Understanding bland altman analysis. *Biochemia medica: Biochemia medica* 25, 2 (2015), 141–151.

[28] Michael Groeneweg, Juan C. Quero, Ilone De Bruijn, Ieneke J C Hartmann, Marie Louise Essink-Bot, Wim C J Hop, and Solko W. Schalm. 1998. Subclinical Hepatic Encephalopathy Impairs Daily Functioning. *Hepatology* 28, 1 (1998), 45–49. https://doi.org/10/dcvrb5

[29] Weixi Gu, Yuxun Zhou, Zimu Zhou, X I Liu, H A N Zou, P E I Zhang, Costas J Spanos, and L I N Zhang. 2017. SugarMate : Non-intrusive Blood Glucose Monitoring with Smartphones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3 (2017), 54:1–27. https://doi.org/10/cn8t

[30] Matthew R Kappus and Jasmohan S Bajaj. 2012. Assessment of Minimal He (With Emphasis on Computerized Psychometric Tests). *Clinical Liver Disease* 16, 1 (2012), 43–55. https://doi.org/10/cm7hkq arXiv:NIHMS150003

[31] Ravi Karkar, Jessica Schroeder, Daniel A. Epstein, Laura R. Pina, Jeffrey Scofield, James Fogarty, Julie A. Kientz, Sean A. Munson, Roger Vilardaga, and Jasmine Zia. 2017. TummyTrials: A Feasibility Study of Using Self-Experimentation to Detect Individualized Food Triggers. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2017)*. 6850–6863. https://doi.org/10/cfbw

[32] Ravi Karkar, Jasmine K. Zia, Roger Vilardaga, Sonali R. Mishra, James Fogarty, Sean A. Munson, and Julie A. Kientz. 2016. A Framework for Self-Experimentation in Personalized Health. *Journal of the American Medical Informatics Association (JAMIA)* 23, 3 (2016), 440–448. https://doi.org/bwtt

[33] Matthew Kay, Dan Morris, MC Schraefel, and Julie A. Kientz. 2013. There's No Such Thing as Gaining a Pound: Reconsidering the Bathroom Scale User Interface. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '13*. ACM Press, New York, New York, USA, 401. https://doi.org/10/bbdw

[34] Logan Kendall, Dan Morris, and Desney Tan. 2015. Blood Pressure Beyond the Clinic: Rethinking a Health Metric for Everyone Logan. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. 1679–1688. https://doi.org/10/ckg7

[35] Gerald Kircheis, Matthias Wettstein, Lars Timmermann, Alfons Schnitzler, and Dieter Häussinger. 2002. Critical Flicker Frequency for Quantification of Low-grade Hepatic Encephalopathy. *Hepatology* 35, 2 (2002), 357–366. https://doi.org/10/dfz3s6

[36] Kenneth D Kochanek, Jiaquan Xu, Sherry L Murphy, Arialdi M Minino, and Hsiang-Ching Kung. 2012. *National Vital Statistics Reports Deaths : Final Data for 2009*. Technical Report 3. Centers for Disease Control and Prevention. 1–117 pages. http://www.cdc.gov/nchs/data/nvsr/nvsr58/nvsr58

[37] Carney Landis. 1951. Something about Flicker-Fusion. *The Scientific Monthly* 73, 5 (1951), 308–314. http://www.jstor.org/stable/20441

[38] Carney Landis. 1954. Determinants of the Critical Flicker-Fusion Threshold. *Physiological Reviews* 32, 2 (1954), 259–286. https://doi.org/10/cjtx

[39] Eric C. Larson, Mayank Goel, Gaetano Boriello, Sonya Heltshe, Margaret Rosenfeld, and Shwetak N. Patel. 2012. SpiroSmart: Using a Microphone to Measure Lung Function on a Mobile Phone. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing - UbiComp '12*. 280–289. https://doi.org/10/b8fz

[40] Ian Li, Anind K. Dey, and Jodi Forlizzi. 2010. A Stage-Based Model of Personal Informatics Systems. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2010)*. New York, New York, USA, 557–566. https://doi.org/bh8zsb

[41] Lena Mamykina, Elizabeth Mynatt, Patricia Davidson, and Daniel Greenblatt. 2008. MAHI: Investigation of Social Scaffolding for Reflective Thinking in Diabetes Management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 477–486. http://doi.acm.org/10/dck23q

[42] Lena Mamykina, Arlene M. Smaldone, and Suzanne R. Bakken. 2015. Adopting the Sensemaking Perspective for Chronic Disease Self-management. *Journal of Biomedical Informatics* 56 (2015), 406–417. https://doi.org/10/f7n6wz arXiv:15334406

[43] Alex Mariakakis, Jacob Baudin, Eric Whitmire, Vardhman Mehta, Megan A Banks, Anthony Law, Lynn McGrath, and Shwetak N Patel. 2017. PupilScreen: Using Smartphones to Assess Traumatic Brain Injury. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 3 (2017), 81. https://doi.org/10/ckg8

[44] Ross A McFarland, A Bertrand Warren, and Charles Karis. 1958. Alterations in Critical Flicker Frequency As a Function of Age and Light: Dark Ratio. *Journal of Experimental Psychology* 56, 6 (1958), 529–538. https://doi.org/10/dbsmvd

[45] Eiman Nabi and Jasmohan S. Bajaj. 2014. Useful Tests for Hepatic Encephalopathy in Clinical Practice. *Current Gastroenterology Reports* 16, 1 (2014), 1–14. https://doi.org/10/f5m6z7 arXiv:NIHMS150003

[46] Rajalakshmi Nandakumar, Shyamnath Gollakota, and Nathaniel F. Watson. 2015. Contactless Sleep Apnea Detection on Smartphones. In *Proceedings of the International Conference on Mobile Systems, Applications, and Services (MobiSys 2015)*. 45–57. https://doi.org/10/b8f6

[47] Raffaele Nardone, Alexandra C Taylor, Yvonne Holler, Francesco Brigo, Piergiorgio Lochner, and Eugen Trinka. 2016. Minimal Hepatic Encephalopathy: A Review. *Neuroscience research* 111 (2016), 1–12. https://doi.org/10/b2bt

[48] F. F. Poordad. 2007. Review Article: The Burden of Hepatic Encephalopathy. *Alimentary Pharmacology and Therapeutics* 25, SUPPL. 1 (2007), 3–9. https://doi.org/10/d5wjgp

[49] Ravi K. Prakash, Thomas A. Brown, and Kevin D. Mullen. 2011. Minimal Hepatic Encephalopathy and Driving: Is the Genie Out of the Bottle? *American Journal of Gastroenterology* 106, 8 (2011), 1415–1416. https://doi.org/10/bsz7qs

[50] Srinivasa Prasad, Radha K. Dhiman, Ajay Duseja, Yogesh K. Chawla, Arpita Sharma, and Ritesh Agarwal. 2007. Lactulose Improves Cognitive Functions and Health-related Quality of Life in Patients With Cirrhosis Who Have Minimal Hepatic Encephalopathy. *Hepatology* 45, 3 (2007), 549–559. https://doi.org/10/b3bgfd

[51] Eva Román, Joan Córdoba, Maria Torrens, Xavier Torras, Càndid Villanueva, Víctor Vargas, Carlos Guarner, and Germán Soriano. 2011. Minimal Hepatic Encephalopathy Is Associated With Falls. *The American Journal of Gastroenterology* 106, 3 (2011), 476–482. https://doi.org/10/c2f9jh

[52] Manuel Romero-Gómez. 2007. Critical Flicker Frequency: It Is Time to Break Down Barriers Surrounding Minimal Hepatic Encephalopathy. *Journal of Hepatology* 47, 1 (2007), 10–11. https://doi.org/10/dbhxk7

[53] Manuel Romero-Gómez, Fernando Boza, Maria Sierra García-Valdecasas, Emilio García, and José Aguilar-Reina. 2001. Subclinical Hepatic Encephalopathy Predicts the Development of Overt Hepatic Encephalopathy. *American Journal of Gastroenterology* 96, 9 (2001), 2718–2723. https://doi.org/10/b3bgfd

[54] Manuel Romero-Gómez, Juan Córdoba, Rodrigo Jover, Juan A. Del Olmo, Marta Ramírez, Ramón Rey, Enrique De Madaria, Carmina Montoliu, David Nuñez, Montse Flavia, Luis Compañy, José M. Rodrigo, and Vicente Felipo. 2007. Value of the Critical Flicker Frequency in Patients With Minimal Hepatic Encephalopathy. *Hepatology* 45, 4 (2007), 879–885. https://doi.org/10/c3cqc4

[55] Steven Scaglione, Stephanie Kliethermes, Guichan Cao, David Shoham, Ramon Durazo, Amy Luke, and Michael L. Volk. 2015. The Epidemiology of Cirrhosis in the United States. *Journal of Clinical Gastroenterology* 49, 8 (2015), 690–696. https://doi.org/10/f7pmv6

[56] Jessica Schroeder, Jane Hoffswell, Chia-Fang Chung, James Fogarty, Sean Munson, and Jasmine Zia. 2017. Supporting Patient-Provider Collaboration to Identify Individual Triggers using Food and Symptom Journals. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17*. 1726–1739. https://doi.org/10/bz28 arXiv:15334406

[57] Praveen Sharma and Barjesh Chander Sharma. 2012. Lactulose for Minimal Hepatic Encephalopathy in Patients With Extrahepatic Portal Vein Obstruction. *Saudi Journal of Gastroenterology* 18, 3 (jun 2012), 168. https://doi.org/10/f3zjzb

[58] Praveen Sharma, Barjesh Chander Sharma, and Shiv Kumar Sarin. 2010. Critical Flicker Frequency for Diagnosis and Assessment of Recovery From Minimal Hepatic Encephalopathy in Patients With Cirrhosis. *Hepatobiliary and Pancreatic Diseases International* 9, 1 (2010), 27–32. https://doi.org/10/b4kbbz

[59] Debbie L. Shawcross, Arthur A. Dunk, Rajiv Jalan, Gerald Kircheis, Robert J. De Knegt, Wim Laleman, John K. Ramage, Heiner Wedemeyer, and Ian E.J. Morgan. 2016. How to Diagnose and Manage Hepatic Encephalopathy: A Consensus Statement on Roles and Responsibilities Beyond the Liver Specialist. *European Journal of Gastroenterology and Hepatology* 28, 2 (2016), 146–152. https://doi.org/10/cjc8

[60] Sandeep Singh Sidhu, Omesh Goyal, Bholeshwar Prashad Mishra, Ajit Sood, Rajoo Singh Chhina, and Ravinder Kumar Soni. 2011. Rifaximin Improves Psychometric Performance and Health-related Quality of Life in Patients With Minimal Hepatic Encephalopathy (the RIME Trial). *American Journal of Gastroenterology* 106, 2 (2011), 307–316. https://doi.org/10/bpv3qd

[61] Xiao Sun, Zongqing Lu, Wenjie Hu, and Guohong Cao. 2015. SymDetector: Detecting Sound-Related Respiratory Symptoms Using Smartphones. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp 2015*. 97–108. https://doi.org/10/ckh6

[62] F. J. Torlot, M. J W McPhail, and S. D. Taylor-Robinson. 2013. Meta-analysis: The Diagnostic Accuracy of Critical Flicker Frequency in Minimal Hepatic Encephalopathy. *Alimentary Pharmacology and Therapeutics* 37, 5 (2013), 527–536. https://doi.org/10/b2bv

[63] Hendrik Vilstrup, Piero Amodio, Jasmohan Bajaj, Juan Cordoba, Peter Ferenci, Kevin D. Mullen, Karin Weissenborn, and Philip Wong. 2014. Hepatic Encephalopathy in Chronic Liver Disease: 2014 Practice Guideline by the American Association for the Study of Liver Diseases and the European Association for the Study of the Liver. *Hepatology* 60, 2 (2014), 715–735. https://doi.org/10/cjc5

[64] Tianzuo Zhan and Wolfgang Stremmel. 2012. The Diagnosis and Treatment of Minimal Hepatic Encephalopathy. *Deutsches Arzteblatt International* 109, 10 (2012), 180–187. https://doi.org/10/cjpb

Received February 2018; revised May 2018; accepted September 2018