Drunk User Interfaces: Final project report

Rafael Caseiro Lemos
University of Copenhagen
Copenhagen, Denmark
qxl764@alumni.ku.dk

Konstantinos Poul Papanikos
University of Copenhagen
Copenhagen, Denmark
xtv647@alumni.ku.dk

Begoña Sustacha Trujillo
University of Copenhagen
Copenhagen, Denmark
frh503@alumni.ku.dk

CHI'20,, April 25–30, 2020, Honolulu, HI, USA
ACM 978-1-4503-6819-3/20/04.
https://doi.org/10.1145/3334480.XXXXXXX

Introduction

Drinking is one of the most common social activities performed when going out with friends. While not inherently dangerous in nature if it is done responsibly, many times people end up drinking excessively, which often leads to bad results, from more minor things such as "drunk texting" all the way to extremely dangerous actions such as "drunk driving", which not only puts the intoxicated person's life at risk, but also other innocent lives which may be involved in an accident through no fault of their own. Statistics show that in 2016, there were 10,497 deaths while driving under the influence of alcohol in the United States of America, which corresponds to 28% of all driving related deaths [7], meaning this is a widespread problem.

Inebriation is considered to be a form of temporary impairment, affecting a person and the way they interact with the environment around them. To detect if a person is inebriated, breathalyzers are primarily used. However, they are mostly used by police enforcement, and they are only used to detect if someone has been driving inebriated, meaning they are not used as a prevention method. In addition, breathalyzers for personal use may be expensive, and their performance varies from model to model [11], meaning they might not be reliable.

To combat this, we will replicate and extend the work done by Mariakakis *et al.* [14], by developing an app that consists of various tasks designed to test the user's psychomotor capabilities. The tasks themselves are based around and/or inspired by methods present in other research papers [2, 16] or used in real-life situations [6].

We extend the original work by reworking the reaction task and by adding two entirely new tasks: a memory task and a walking task. Based on the user's performance and smart phone sensor readings, we will calculate the state of the Blood Alcohol Level (BAL) with the help of a machine learning model and inform the user if he has a safe value of BAL or not, with the safe value being at most the maximum limit of alcohol allowed to drive in his country of residence minus 0.01%, and the unsafe value being the maximum limit of alcohol allowed to drive in his country of residence plus 0.01%. Values between these two are considered inconclusive, since they are very close to the limit and thus it would be dangerous to give a safe or unsafe BAL state since minor details can impact the results. This app will help prevent situations like the ones previously mentioned. This will be validated by performing a user study, which will compare the results given by the app with a breathalyzer.

The primary contributions of this work are:

- An extension of the DUI app presented in the work done by Mariakakis *et al.* [14], by reworking the reaction task and by introducing two new tasks that can be used to estimate the BAL of a user; more specifically, a memory and a walking task.
- A machine learning model that converts the data logged in the app to a BAL state
- A user study that demonstrated the app's ability to correctly identify the BAL state of a user, using a breathalyzer as a baseline.

We collected data from 10 participants to train our model and evaluated the app witha a further 5, in which the participants used the DUI app multiple times in order to predict their BAL state. Results show that our app had 63.3% accuracy in predicting the correct BAL state.

Related Work

Our work is a direct extension of the work by Mariakakis *et al.* [14]. We highlight and mention some of the methods used to detect inebriation, previous research that explores apps that can detect if the user is inebriated and the paper we are extending.

Detecting and Measuring BAL

There are numerous ways to detect and measure the BAL level of a person. There are simple, proven methods, like drawing a blood sample and analyzing its contents, and more experimental ones, such as measuring BAL using photoplethysmography (PPG). PPG allows detection of certain properties and changes in our blood [4], by shining an LED on the skin. Since blood has slight color changes when alcohol is present, we can detect the presence of it using PPG.

The most commonly used way to detect and measure BAL, however, is with the use of a breathalyzer. Breathalyzers are the *de facto* method used by law enforcement to measure BAL, but personal use breathalyzers are also available for sale to the general public. These vary a lot in price, ranging from 15$ to 130$ on Amazon. However, the different price points also reflect the quality and accuracy of the breathalyzer. A previous study [11] compared different personal use breathalyzer and found out that the cheapest one, costing 2.99£, only identified 26.3% of all cases in which the BAL was above the legal limit, while the most expensive one, valued at 299£, identified 94.7% of all cases in which the BAL was above the legal limit.

The problem with the aforementioned methods is that not everyone has access to such devices, and investing in them means that people will have something that will only be used for a specific function only.

In case none of the above is available, a common technique used by law enforcement to assess inebriation is the Field Sobriety Test [6]. It consists of three tests designed to assess balance, coordination and the capbility to focus on more than one thing at the same time, and it has been scientifically proven to correctly detect that a person is inebriated. However, there is no way to determine what is the BAL value of a person using a technique like this.

Detecting Alcohol Consumption Using Smartphones

There has been some work done in which a smartphones' capabilities were used in order to detect if a user has consumed alcohol.

Arnold *et al.* investigated if the alcohol intoxication level of a smartphone user could be inferred from their gait[10]. The term gait refers to the pattern of movement of our limbs during movement on a solid surface. Since alcohol impacts a person's ability to walk, analyzing the gait could give insight into if a person was inebriated. To test this, the authors collected gait data from a group of drinks, by using the accelerometer data from their smartphones. With the data, they used a random forest classifier in order to identify the number of drinks the user had consumed. All of this was then used to develop an app called *AlcoGait*, that can detect the number of drinks the user has consumed by taking their gait readings from the last hour and comparing them to the ones captured during the previous 24 hours to detect any noticeable changes. This work was later extended by Nassi *et al.*, in which wearable devices were used instead of a smartphone [15].

Dai *et al.* developed an app that detects if the user is driving inebriated [13]. They achieved this by computing acceleration values from the smartphone's accelerometer and orientation sensors, and comparing it with typical drunk driving patterns extracted from real driving tests. If it drunk driving is detected, the app will alert the driver, or call the police automatically to intercept him.

Bae *et al.* ran an experiment with an app in which they used the built-in smartphone sensors and a daily survey to monitor users drinking activities for 28 days [12] and with the collected data developed a machine-learning model that would help identify non-drinking, drinking and heavy drinking episodes. Their results show that mobile phone sensors can be used for automated and continuous monitoring of drinking episodes and alcohol consumption.

Lastly, Willoughby *et al.* explored if alcohol intoxication could be detected by analyzing selfie images [17]. Their findings reveal that facial lines changed significantly after a person has consumed alcohol. A machine learning model was developed based on this, leading to the creation of the *DrunkSelfie* app, which estimates a person's drunkenness state from a selfie.

Original Drunk User Interfaces paper

Mariakakis *et al.* developed the original work [14] we are extending. Their aim was to develop an app that was able to calculate the BAL of a user using their smartphone, which could be used as a substitute for a breathalyzer, by combining different methods from the HCI literature and medical communities that evaluated inebriation into one single place.

The end result was *DUI*, an app comprised of five tasks that assess a person's motor coordination and cognition:

- A typing task, intended to measure motor coordination while texting;
- A swiping task, intended to measure fine motor control through gesturing;
- A balancing+heart rate task, intended to measure heart rate and coordination;
- A simple reaction task, intended to capture alertness and motor speed;
- A choice reaction task, with the same as the above task.

From these tasks, the authors logged 51 human performance metrics and sensor data that were used to train a machine learning model; more specifically, the app uses random forest regression models. This model takes into account user-specific learning, and is able to estimate a person's BAL.

The authors also conducted a user study by comparing the results from *DUI* with a breathalyzer. They observed that DUI cold estimate a person's BAL with an absolute mean error of $0.005\% \pm 0.007\%$ and a Pearson's correlation coefficient of 0.96 with breathalyzer measurements.

Drunk User Interfaces

In our work, we extend the work by Mariakakis *et al.* [14], by adding two completely new tasks that test aspects commonly affected by alcohol, such as memory and balance while walking, not covered in the original work, while also reworking one of the existing ones. Taking inspiration from some of the related work, we also implement a pre-trained model instead of a learning based one, so that the user can use the app and get results immediately, without having to calibrate it first.

App

The DUI app will help users without access to a breathalyzer assess if they have a safe level of alcohol in their blood. This is done by conducting 3 tasks which will test the user's reactions, memory and balance. During the tasks, various metrics will be logged, which will be used to estimate and inform the user if he has a safe BAL level. It is a native Android APP, developed using Android Studio and written in Kotlin.

When the app is first opened, a welcome page is displayed, where the user is welcomed and is explained what the app is about. Once ready, the user is can click the start button, which will redirect him to the instructions page of the first task.

Before the start of every task, an instruction screen is shown, where the user is informed on how the task should be performed, pressing the start button once they are ready to begin. Once the task is finished, they are shown another interface indicating that the task is completed, and a button to move to the next one. The phone also vibrates when the task starts and ends, in order to give the user an additional form of feedback.

The first task is a *Reaction Task*, during which the user will be presented with a square, positioned in a random part of the screen. The goal of this task is to make the user tap on the square as fast as possible. The user does this 5 times, with each time the square being in a different position in the screen, and a 1 second pause between clicking a square and the next one showing up. This is a form of what is called a Simple Reaction Time Task [2], where users make a response when a stimuli appears, which is used when basic processes such as perception and response time want to be measured, which is the aim of this task. There are two measurements taken from this task: the average reaction time across all 5 tries, with reaction time being defined as the time the user takes to click the square from when it first appears on screen, and the number of misclicks a user does while trying to click the square, with a misclick being any area of the screen other than the square.

The second task is a *Memory Task*. The task's main objective is to evaluate the short-term memory capability of the user. Studies have shown that alcohol has an effect on a person's working memory [16], meaning it is an appropriate way of evaluating sobriety. The task consists in a sequence of five random numbers between 0 and 9 being displayed one at a time in one second intervals. Since two numbers in a row can be the same, the phone vibrates when each number of the sequence shows up, in order to let the user know that a new number in the sequence is being shown. Once the whole sequence has been shown, the user is asked to input it correctly. This process is repeated for another two times, with different sequences of numbers every time. This was inspired by the "Auditory Array and Sequence Tasks" from [16], replacing audio stimuli (voice that called out the numbers) to visual stimuli (numbers displayed on screen), and increasing the time between numbers from half a second to 1 second, since it was too fast and challenging even for a sober person. The data extracted from this task is how many numbers they inserted incorrectly, as well as the average time the user took to input the sequence of numbers across the 3 tries. The third and last task is a *Walking Task*, which consists in the user walking for five seconds, while having the phone laying on the palm of their hand, parallel to the floor, with their arm outstretched, trying to keep it as balanced as possible. This is inspired by the "walk-and-turn" task from the Field Sobriety Test, one of the most used and successful methods to detect if someone is inebriated [6]. Using the phone's accelerometer and magnetometer, we calculate the phone's orientation in degrees in relation to the X-Axis (pitch) and the Y-Axis (roll). These values are registered during the 5 seconds the user is walking. Once the task is done, the standard deviation from all the values in each one of the axis is calculated, allowing us to obtain how much degrees the phone deviated in both the X and Y axis during movement. Lower degree values indicate the user was more stable while walking, and the opposite applies for higher values. These two values will be used for the BAL estimation.

When the last task is completed, the user is taken to the results screen, in which they are shown how they they performed in their tasks, as well as displaying their BAL state:

- If the BAL is 0.04% or below, "SAFE" is displayed;
- If the BAL is between 0.04% and 0.06%, "INCONCLUSIVE" is displayed;
- If the BAL is 0.06% or above, "UNSAFE" is displayed.

Machine Learning

Each of the mentioned 3 tasks of our App generates a set of human metrics that are used as features for training in a Random Forest model, which will then predict one of by estimating 3 possible BAL states, which are **"Safe" , "Inconclusive"** and **"Unsafe "**. Measuring participants' performance metrics across the different trials and BAL values

Task	Features
Reaction Task	Average Reaction Time Misclicks
Memory Task	Average Input Time Incorrect Numbers
Walking Task	X-Axis Deviation Y-Axis Deviation

Table 1: The features extracted from the app.

was of great interest, since that way we could observe what metrics get more affected by the consumption of alcohol, allowing us to see which task got the most affected by alcohol consumption.

From the 3 tasks that were performed, 6 features were extracted (Table 1) and used for training. The model was built in Python using the scikit-learn package, and then exported to our app using the sklearn-porter package. Due to the low number of data points for training (60 in total), overfitting was a concern. This was verified by using the default number of trees, 100, which did not work well with our data. We settled on using 10 trees, as during our tests it was the configuration that consistently gave the best results with the data we had. All features were use to predict the final state and no random state was set so the predictions stayed consistent.

We were also able to extract the more important features by assigning them a score input features based on their contribution for predicting a BAL state. The walking task features both contribute to almost 70% , the memory task features contributed to 18% and lastly the reaction task features contributed 12%, signalling that, in our training data, the walking task was the one that suffered the biggest changes with the increase in inebriation.

User Study

To collect data and evaluate the performance of our app, we conducted a user study in which we analyze the participants' performance on the metrics that are logged and compare the expected BAL state with the one given by the app, across six increasing BAL levels. We hypothesized that the participants' performance would gradually worsen the higher the BAL level got.

Participants

Fifteen participants, of which 11 were male and 4 female, took part in our experiment. The average age across all participants was approximately 23 years, while the average weight was 70.4 kg. All participants used smartphones daily and had extensive experience with mobile apps.

Design

The conducted experiment featured one Independent Variable, *BAL Value*, containing 6 different levels, with each level being an increasingly higher value of BAL, starting at 0.00% for the first level, and ending at around 0.1% for the last level. Since all participants used the app in all six different BAL levels, the experiment followed a within-subject design.

The BAL limit for driving in Denmark is 0.05%; however, we ran the experiment until BAL values of around 0.1%. This was due to:

- The model being pre-trained, which meant that, for data collection, it would be beneficial to obtain performance metrics across various different BAL values, including ones that were above the legal limit, which would have the app have better predictions;
- At around 0.1% BAL, "reaction time and control will be reduced, speech will be slurred, thinking and reasoning are slower, and the ability to coordinate your arms and legs is poor" [5]. Most of these impairments are something that the tasks in the app directly test, so comparing the results at 0.00% BAL to 0.1% will be a good indicator that the tasks are designed correctly and working as intended.

In order to achieve the final desired BAL value, 45 ml shots of 40% volume vodka were administered. Following this configuration, a shot is estimated to increase a person's BAL value by approximately 0.02% [8]. This allowed for a steady and safe increase of a participant's BAL value, and allowed us to observe how these changes affected the performance metrics that the app logs.

To obtain the BAL value of a participant, we took the average value given by the breathalyzer from 3 uses. This was done due to the low quality of the breathalyzer and to minimize erroneous BAL values. Even though the legal Danish BAL limit is 0.05%, values between 0.04% and 0.06% were classified as inconclusive, since minor details can make the difference between being legal or not, meaning it is a sensitive area that we decided to not give a "correct" label.

In each of the BAL values, the app registered 6 performance metrics, meaning the experiment contained 6 dependent variables. These were:

- Average reaction time, defined as the average time that a user takes to click the square from the moment it first shows up across all 5 repetitions in the reaction task;
- Missclicks, defined as the number of times the user fails to click the square during the reaction task;
- Average decision time, defined as the average time that a user takes to input the number sequence from the moment the input field first shows up until he confirms his input across all 3 repetitions in the memory task;
- Incorrect numbers, defined as the numbers that the user input incorrectly during the memory task;
- X-Axis deviation, defined as the number of degrees the phone varied during movement in relation to the X-Axis (pitch) during the walking task;
- Y-Axis deviation, defined as the number of degrees the phone varied during movement in relation to the Y-Axis (roll) during the walking task.

Only one participant at a time took part in the experiment, which was conducted in a quiet environment, in order to minimize distractions and keep the participants focused, ensuring that we got representative results to train the model with. All participants were given time to test the app to get used to it, in order to minimize learnability effects. Out of the 15 participants, the results of the first 10 were used to train the model, while the last 5 participants were used to evaluate the app. The experiment took around 2 hours per participant, and was conducted mid to late afternoon, to avoid fatigue in the participants, which causes similar impairments to inebriation.

Setup

The device used in the experiment was a Huawei P30 Pro smartphone, possessing an 6.27 inch screen with 1080x2340 pixels and running Android 11.0.0 as the Operating System. The experiment app was pre-loaded in the device. The built-in screen capture software was used to record the screen. The app results were manually registered on a laptop.

To measure the BAL level of the participants, we used the "Digital Breath Alcohol Tester" breathalyzer [9]. It features a measuring range from 0.00% to 0.19% BAL, and has an accuracy of $\pm 0.01\%$ BAL. Absolut Blue Vodka 40% was used as the drink, served on a 50 ml shot glass.

The experiment was conducted in a quiet, indoors environment, with only the experimenters present in addition to the participant, in order to minimize distractions. The environment contained a hallway that was long enough in order for the participants to complete the walking task successfully in a straight line. For the reaction and memory tasks, the participants sat in front of a desk.

Procedure

Due to the nature of the experiment, potential participants had to satisfy various requirements to ensure their safety and follow a set of guidelines we developed so that the experiment could be carried out as planned. In order to participate, they would have to be at least 18 years old, which is the minimum legal age for drinking spirits in Denmark, as well as confirm that they had no alcoholism problems or were on medication that could interact with alcohol. In addition, female participants also had to prove that they were not pregnant. Our own guidelines required that users did not excessively consume alcoholic beverages 24 hours prior to the experiment, that they did not eat anything 3 hours prior to the experiment and that they had a valid *Coronapas*. Participants that satisfied the requirements and accepted the guidelines were invited to take part in the experiment. Before each participant arrived, the experimenters tested the app to make sure everything was running as intended, and disinfected the breathalyzer's mouthpiece, as well as the shot glass for hygienic and safety purposes. Afterwards, the participants would be welcomed and was then given a form that explained the purpose of the experiment as well as some additional details. They would then fill in their name, age, gender and weight, as well as confirm that they followed all requirements and guidelines, that they consented to his data being collected and used for our experiment and that they took responsibility for any actions taken while inebriated. If they consented to all points, the experiment would progress.

The participants was then introduced to the DUI app by one of the experimenters. He explained how it worked, what were the tasks and how to perform them, and what were the metrics that influenced the final result. He experimenter then gave a short demonstration of the app in action. The participants then used the app for 5 times in order to get familiarized with it. They were free to use the smartphone as they saw fit, as long as it was in their hands.

Once the training period was over, one of the experimenters explained the BAL measuring process, as well as the alcohol consumption process. The participants were informed that they were free to stop at any time in case they were feeling sick or uncomfortable.

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6
ART	0.503	0.492	0.508	0.505	0.525	0.556
M	0.200	0.133	0.000	0.066	0.066	0.466
AIT	4.144	3.828	3.993	4.046	4.257	4.205
IN	1.466	1.400	0.866	1.200	1.666	1.733
XAD	4.002	4.525	4.278	4.815	4.904	6.348
YAD	4.169	4.546	4.422	4.300	5.385	6.214
BAL	0.000	0.029	0.042	0.059	0.079	0.102

Table 2: Means of each feature during each trial. ART = Avg. Reaction Time , M = Misclicks , AIT = Avg. Input Time , IN = Incorrect Numbers , XAD = X-Axis Deviation , YAD = Y-Axis Deviation , BAL = Blood Alcohol Level

Before using the app, the participants used the Breathalyzer 3 times to measure their BAL, with the average value across all 3 uses being their current BAL. Before using the breathalyzer, the participants rinsed their mouth with water, in order to remove any possible alcohol residues. If the initial BAL value was not 0.00%, the experiment would not progress. The participants then used the app until all tasks were completed, with the results being registered by one of the experimenters. Afterwards, they would drink a shot. Once 15-20 minutes had passed, which is the estimated time it takes for the blood to absorb the alcohol, the participants would use the Breathalyzer again. This process would be repeated until 5 shots had been consumed, resulting in a total of 6 app usages.

Once the experiment was finished, food was given to the participants in order to help reduce inebriation effects. The participants were free to leave when they felt safe, and they were thanked for their time and collaboration.

Results

In this section, we describe and analyze the results of our user study.

Table 2 shows the mean values of each feature of the 15 participants in each trial, plus the average BAL value for each one of the trials.

As a general rule, results improved or fluctuate from a sober state up until the third or fourth trial, at which point the results started to gradually get worse. The only metric in which the average result measured in the most intoxicated state, corresponding to trial 6, was not the worst was AIT (Average Input Time); all others had their worst result in this trial, as expected.

None of the performance metrics got consistently worse with each trial as hypothesized. The metrics that got the most affected the higher the BAL value were the ones related to the walking task, XAD (X-Axis Deviation) and YAD (Y-Axis Deviation).

The BAL increased by 0.02% on average, the expected amount for a shot of vodka. The final BAL value was 0.102%, very similar to the target amount we aimed to achieve. Female participants had an average BAL value of 0.12% at the end of the experiment, while male participants had an average BAL value of 0.09% at the end of the experiment.

Label	Accuracy
SAFE	83%
INCONCLUSIVE	37.5%
UNSAFE	60%

Table 3: Prediction accuracy for each BAL state.

For the 5 participants that took part in the evaluation, the app correctly identified 63.3% of the expected BAL states. **Table 3** shows the accuracy percentages for each of the three BAL states. The **"SAFE"** BAL state was the most correctly identified one, with an accuracy value of 83%. The **"UNSAFE"** label followed, having been correctly identified in 60% of all cases. The lowest accuracy was achieved on the **"INCONCLUSIVE"** BAL state, with only 37.5% of the cases being correctly identified. The app correctly identified the BAL state on 100% of all cases in which the participants were completely sober (BAL value of 0.00%), as well as when they were at their highest BAL value.

Discussion

Analyzing the results, the app performed better than what we expected, predicting correctly almost two thirds of all measured BAL states. Due to the fact that we had a small dataset to train the model with, coupled with the fact that people have wildly different ways of interacting with smartphones, bad results were expected. However, this was not the case. One of the reasons for this might be due to the fact that all the participants were experienced with smartphones and apps. This experience made it easier for them to adapt to the app and use it, meaning that the performance metrics of all participants, while different, do not vary as much between them as we would have expected. However, the performance metrics for people with little experience or impairments would be much worse, and as a result the model would always classify them as **"UNSAFE"** even though they might not be, due to the fact that the model was trained with data from experienced users.

Some of BAL states were more easily identified than others, particularly the **"SAFE"** BAL state. This is due to the fact that most of the data points present in the dataset are from safe BAL values. This means that more data is present for this state, and therefore the model can more easily identify thee cases. The opposite occurs for the **"INCONCLUSIVE"** BAL state, which only corresponded to 9 data points out of 60 in the training data set, meaning it was much harder to correctly predict. In these cases, the app would predict mostly either safe or unsafe, meaning that this particular state, added to avoid incorrect prediction when the BAL value was very close to 0.05%, ended up doing exactly what we were trying to avoid and causing confusion. If this state waws not implemented, more data would available to correctly classify safe or unsafe states, which could have made the model more accurate overall. However, the app correctly predict all cases in which the user was completely sober or had an high BAL value, meaning our app would be viable in these specific situations.

A big limitation in this project was the quality of the breathalyzer that was used. Being one of the cheapest on the market, its accuracy left a lot to be desired, and this meant that we were potentially classifying BAL states incorrectly when building our model. While an attempt was made to mitigate this by having the participants use the breathalyzer multiple times to avoid erroneous values, some situations still arose in which the value did not correspond to what was expected after the amount of shots they had consumed.

Another limitation that we believe had a big effect on the app results was the learnability aspect associated with it. Since we did not use a learning-based approach for our model, we had to avoid that the participants performed better the more times we used the app. We let them use the app 5 times before the experiment started so that the impacts of this were minimized. However, it did not seem to have an effect, since the majority of participants improved their results on during their first three usages, even though their BAL was increasing. This was not ideal at all for the model, as it is expected that the performance of the users to decrease the more inebriated they get, which did not happen. Only starting with the fourth usage did results start to get progressively worse, possibly due to the fact that the effect of alcohol was starting to have a noticeable effect. Since we are only predicting states, the impact of this was not as noticeable in comparison to if we had tried to predict the exact value.

The participants were asked about their weight and gender. As mentioned in the results sections, female participants got a higher BAL in comparison with the male participants, when consuming the same amount of alcohol. This lead to them generally performing worse the more inebriated they got, comparatively to male participants. This can be explained by the National Institute on Alcohol Abuse and Alcoholism, which explains how *"women absorb and metabolize alcohol differently than men. In general, women have less body water than men of similar body weight, so that women achieve higher concentrations of alcohol in the blood after drinking equivalent amounts of alcohol"*[3]. Moreover, the average weight of the female participants was considerably lower in comparison to the average weight of the male participants, which has been proved to have a big impact on the effects of alcohol. Stanford University ranks Biological Sex and Weight as the first two factors that affect how alcohol is absorbed and metabolized[1].

The experiment environment might also not have been the most representative one. The participants tried the app on quiet environment with only the experimenters alongside. Although there could have been some sound distraction from outside noises, they had the opportunity to concentrate while they were performing the tasks. We believe this to be not the real context the users will be using the app on. If a user wants to know their sober state, it is likely they already had something to drink, which is normally done in a social scenario with either more people or music, meaning they would have plenty of things to serve as distractions, considerably diminishing their performance. Therefore, we can argue that the context our experiment was conducted at was not completely realistic. However, we consider it was the best solution to guarantee representative and valid results, since it was conducted in a controlled environment meaning it was easier to enforce the same conditions to all participants.

In regards to the tasks themselves, some were found, especially in the walking task. Even though it was the most representative task to detect inebriation, some factors might have influenced the results. First of all, a problem that kept on happening was the fact that users clicked to start the tasked accidentally without getting in position. This meant that they would have to rush to the starting position once they realized it, often starting the task already in movement, which meant that some values for this task were not representative. Also, while the task should be performed by placing the phone on the hand with the arm outstretched, people might have different ways of positioning their arm (different heights, for example), and things like arm length and user height might also be factors that influence the results for this task, not only the alcohol level. Also, certain features seemed to barely have been affected by the increased BAL values, such as the number of misclicks in the reaction tasks, so other features should be taken into consideration.

Future work on this app should focus on improving the model. Due to the variability of factors that might influence the results, such as experience or impairments, a pretrained model is not the best solution, since even using an extremely large dataset that tried to cover a lot of possible cases to train the model might not be enough due to the sheer different number of ways how people interact with smartphones. Another type of machine learning model could be tested, such as K-Nearest Neighbors, but we believe that the approach taken in the original work by Mariakakis *et al.* [14], which learns from the user in order to then know how to predict, is the best approach, since it can be adapted to all individuals. Exploring other types of tasks would also give more insight into which ones work best for detecting inebriation. As an example, in the memory task, instead of users having to memorize a sequence by looking at the numbers, they might have to memorize the sequence based on audio output only, without any visual cues. Also, it would be worth exploring how conducting the experiment in a more realistic environment would affect the results and the accuracy of the app.

Conclusion

Drinking is a fun activity to do with friends, but it can be dangerous if not done with moderation, and it is the cause of many driving related deaths. Available methods are either expensive, unreliable or not practical, meaning it is not easy to inform the user about they current sobriety state. To combat this, we developed DUI, an app, that can help users estimate their BAL state after alcohol consumption, by gathering performance metrics obtained by completing a reaction, a memory and a walking task and running them through a machine learning model. Our user study showed that it can correctly predict 63.3% of BAL states.

Video Link

https://www.youtube.com/watch?v=uvr5yRLVqXc

REFERENCES

[1] Factors That Affect How Alcohol is Absorbed Metabolized. (????). https://alcohol.stanford.edu/alcohol-drug-info/buzz-buzz/factors-affect-how-alcohol-absorbed

[2] Simple Reaction Time Task. (????). https://scienceofbehaviorchange.org/measures/simple-reaction-time-task/

[3] 1999. Are Women More Vulnerable to Alcohol's Effects? (1999). https://pubs.niaaa.nih.gov/publications/aa46.htm

[4] 2019. Photoplethysmography. (2019). https://www.news-medical.net/health/Photoplethysmography-(PPG).aspx

[5] 2020. Blood Alcohol Level Effects on the Body. (2020). https://www.alcohol.org/effects/blood-alcohol-concentration/

[6] 2020. Field Sobriety Test to Assess Drunk Driving. (2020). https://www.verywellmind.com/field-sobriety-test-67159

[7] 2020. Impaired Driving: Get the Facts. (2020). https://www.cdc.gov/transportationsafety/impaired_driving/impaired-drv_factsheet.html

[8] 2021. Calcoholator. (2021). https://www.educalcool.qc.ca/en/tools/calcoholator/

[9] 2021. Digital Breath Alcohol Tester. (2021). https://www.studyshop.dk/alkometer-alkoholtester?gclid=CjwKCAjwhMmEBhBwEiwAXwFoEeIlg6pKHJZ335Rgbt0Rnu8bw_hwv9aXdXpuht8N7lxkfhNS_3m3ZBoCw4sQAvD_BwE

[10] Zachary Arnold, Danielle Larose, and Emmanuel Agu. 2015. Smartphone Inference of Alcohol Consumption Levels from Gait. In *2015 International Conference on Healthcare Informatics*. 417–426. DOI: http://dx.doi.org/10.1109/ICHI.2015.59

[11] Helen Ashdown, Susannah Fleming, Elizabeth Spencer, Matthew Thompson, and Richard Stevens. 2014. Diagnostic accuracy study of three alcohol breathalysers marketed for sale to the public. *BMJ open* 4 (12 2014), e005811. DOI: http://dx.doi.org/10.1136/bmjopen-2014-005811

[12] Sangwon Bae, Denzil Ferreira, Brian Suffoletto, Juan C. Puyana, Ryan Kurtz, Tammy Chung, and Anind K. Dey. 2017. Detecting Drinking Episodes in Young Adults Using Smartphone-Based Sensors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 2, Article 5 (June 2017), 36 pages. DOI: http://dx.doi.org/10.1145/3090051

[13] Jiangpeng Dai, Jin Teng, Xiaole Bai, Zhaohui Shen, and Dong Xuan. 2010. Mobile phone based drunk driving detection. In *2010 4th International Conference on Pervasive Computing Technologies for Healthcare*. 1–8. DOI: http://dx.doi.org/10.4108/ICST.PERVASIVEHEALTH2010.8901

[14] Alex Mariakakis, Sayna Parsi, Shwetak N. Patel, and Jacob O. Wobbrock. 2018. Drunk User Interfaces: Determining Blood Alcohol Level through Everyday Smartphone Tasks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. DOI: http://dx.doi.org/10.1145/3173574.3173808

[15] Ben Nassi, Lior Rokach, and Yuval Elovici. The Age of Testifying Wearable Devices: The Case of Intoxication Detection. (????).

[16] J Scott Saults, Nelson Cowan, Kenneth J Sher, and Matthew V Moreno. 2007. Differential effects of alcohol on working memory: distinguishing multiple processes. *Experimental and Clinical psychopharmacology* 15, 6 (2007), 576.

[17] Colin Willoughby, Ian Banatoski, Paul Roberts, and Emmanuel Agu. 2019. DrunkSelfie: Intoxication Detection from Smartphone Facial Images. In *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, Vol. 2. 496–501. DOI: http://dx.doi.org/10.1109/COMPSAC.2019.10255