ADAPTIVE SHOOTING ASSISTANT

Final Project DGMD S-14

Gurleen Bath, Brendan Keegan, DianaLiu,

Ben Trey, Sishir Yeety

Wearable Devices and Computer Vision | Treety Bio

Table of Contents

Introduction

Players in the NBA are often lauded as the best basketball players in the world. They combine a rare set of genetics and elite skills to perform at the highest level. The abilities of these athletes appear to give them a clear advantage over amature athletes.

However, if you want to know the most free throws hit in a row it's 5221 set by Ted St. Martin. At the age of 55, a former high school basketball player that specialized in defense set the record. [1] If you want to know the runner up it's Thomas Amberry with 2,750. Amberry, a podiatrist, set the record at the age of 71, with the help of several of the Chicago Bulls training staff. [2] In fact you won't even see a professional basketball player until the 6th position where Harrold "Bunny" Levitt sneaks in with 499 freethrows, a record set in 1935 while he traveled with the Harlem Globetrotters. [3] At the three point line you will find a similar lack of professionals with the record being held by Al Callejas. [4] If you want to compare the record holders to NBA players, the most consecutive free throws during games is 97 by Michael Williams [5] and the most consecutive three pointers in game is 14 by Steph Curry. [6] Although these numbers are not exactly comparable, if you calculate the percentage to achieve St. Martin's achievement one out of 5000 tries a shooter would need to make 9% more free throws than any measured NBA player.

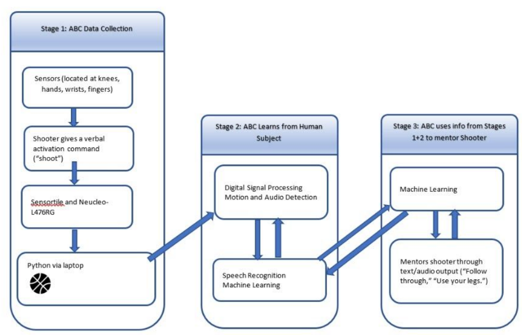
It seems when you say NBA players have "a rare set of genetics and elite skills", they have exactly that. They often are often great athletes that chose to play basketball. When compared to the average Joe they often do not always meet expectations. Injuries, size, and limited mobility will often inhibit their performance.

When looking at shooting, the mechanics of shooting a basketball have been optimized and are well understood. A coach can tell you the motion and position of every body part throughout the shot. A problem arises when taller players, or players with a reduced range of motion try to replicate this form. In fact most players will adapt their motion to accommodate their body. For instance Kobe Bryant didn't point his feet at the hoop to allow for the range of motion of his elbow, Shaq can't bend his wrist due to a childhood injury, Durant can't place the ball at a point on his forehead while maintaining a reasonable bend of his elbow, and even Steph Curry places the ball below his chin as he starts his motion. Historically great shooters such as Peja Stojakovic, and Larry Bird incorporated the off hand, and the only suggestion a coach could give them is to stop using their off hand. With this in mind a coach should assist a shooter by understanding their muscle control, stature, and quirks.

This project aims to give personalized feedback to a shooter. By grouping similar shots together from historic data a model will instruct the shooter using this group. By looping through the historic effect of suggestions on the shooter the model will select the best suggestion from this list.

Theoretical Design

In the design of the apparatus the physics should primarily concern the designers. The system of parabolic motion given an initial velocity is thoroughly studied. [7] The system itself with the uncertainty should give better results than the usual data science methods. Classically given an initial position and velocity the Uniqueness Theorem for Second Order Linear Differential Equations should give an absolute answer if the shooter made the shot. A complication arises when the measured variables do not meet the standard inputs of the problem. A wearable device may not easily measure initial position and velocity. In situations with the measurement of particular features, the specific wearable technology will refine the design.



Work Flow

Our device will rely on the ST Sensor Tile as the central accelerometer and gyroscope. The ST Sensor Tile makes a perfect choice for this project given its libraries [8] and community. [9]

Given the infrastructure of the Sensor Tile the measured values include sound, pressure, temperature, direction, acceleration, and angular speed. For our particular application acceleration and angular velocity will influence the final trajectory the most. For a particular interval of data the computationally inexpensive features include, the mean, max, min, and integration with respect to time. As the device can't determine the initial position prior to shooting, or the position of the ball at the release, the computationally inexpensive variables may replace a full treatment of the problem.

While the firmware on the tile will give full access to the measurements through an USB connection this particular application requires a wireless connection. ST gave a great solution to data transfer through the tile cradle, battery, and BLE connection. [10] To allow for measurement of the data ST also published the blue_st_sdk library for Python. They also publish a similar library for Java which could make the device cross platform scalable for Android devices. There does still exist a problem for Windows based systems. Microsoft does not easily allow for access to their BlueTooth connections. As our device is a prototype we may develop our device in the Linux environment and leave these details for later iterations. There does remain one problem with using the blue_st_sdk library for Python because it prints the data on the screen instead of offering a method to set a variable to this value. Currently overwriting the print method to record the data to a global variable temporarily solves this issue. A further improvement may include altering the blue_st_sdk library.

Initially we decided several locations to place sensors on the body. Some ideal locations would include the knee, elbow, wrist, and back of the hand. Obviously extra sensors would improve accuracy but a trade off exists where after a certain point additional sensors will not warrant their inclusion in the design. The primary problem remains the symmetry about the axis perpendicular to the ground. The device relies heavily on the shooter aligning themselves with the basket. Given this consideration we decided to place one sensor on the outside of the hand. This placement will encompass several motions in the shot such as, the jump, extension of the elbow and snap of the wrist. The ST cradle makes an excellent enclosure for this placement.

Upon capturing the data the device will need to determine how far the distance between the shooter and the hoop. Given the readily measured and calculated values the device will need to group like shots based on the measured values but also on the result of the shot. A great out of the box method for grouping these shots is the KMeans module from the sklean learn library.

After finding similar shots the device can use an appropriate model to predict the result of giving a particular suggestion.

Once the sklearn library finds the relevant data it fits a Logistic Model to the chosen data set. As making or missing a shot gives a boolean outcome the Logistic Model seems like the most reasonable model from sklearn. We also considered a Random Forrest model with limited success. The Random Forest Model may become more appealing if we add additional outcomes to the result such as missing, left, right, short, or long. The model may also increase in accuracy if the selection of features is automated to reduce the dimensionality of the problem.

Finally the device needs to choose the best suggestion for the shooter. As discussed earlier the variations of human anatomy will give different measurements from the Sensor Tile given each suggestion. As KMeans creates several different groups of shots from different distances, a ratio of the measured values before and after the suggestion will remove the dependence on one cluster. After the creation of the model the device could cycle through the possible suggestions and their average effect on previous attempts, put them in the model, and determine which suggestion will give the best result.

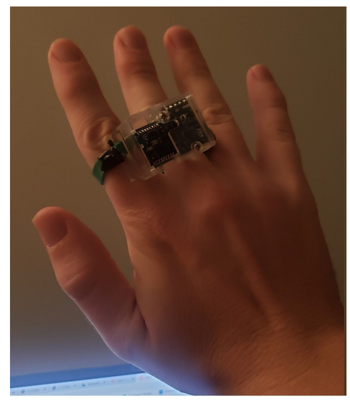
Given this flow the shooter will need to give several pieces of information outside of the measurements from the Sensor Tile to operate the device. First they will need to tell the device when to start recording data. Secondly the user will need to tell the device if they are shooting from a new spot or an old spot. This will aid in the measurement of the mean ratio of previous value to new value given a suggestion. In this prototype the user will also choose the BLE device to connect to. In later iterations the device may contain a hard coding of the MAC address to avoid this prompt. In addition to giving input the user should also receive input out of the speaker or possibly headphones.

Given the large computational expenditure of the sklearn library a lag will exist between the shooter finishing a shot and the device giving the best suggestion. The code may mitigate this by running threads that separate the data using KMeans, and getting a model for each shot prior to the shooter shooting. This will make the operation feel seamless to the user.

Although threads may remove the lag, the user will still need to record data prior to effective operation. Given a 20 groups of shots from varying distances and a Logistic Model with 5 parameters this means the user may need to shoot up to 100 shots before the model starts making predictions. In other devices training the model before the user encounters it overcomes this problem. For this situation the individual characteristics of the user play a tremendous importance. In this iteration we will also determine the effectiveness of creating data based on early measurements.

Model Implementation

The first test of the device was to take data while wearing it on the outside of the hand. The ST Cradle housed the device and two rubber straps attached the device to the first two fingers. The data collection consisted of two separate events. The first event tested the ability of KMeans to partition the data into similar shots. We shot 5 times at varying distances and moved a foot back after each set of shots. This gives an ordered set of shots with increasing distance. The second set of data consisted of 100 free throws. This gives a data set to analyze the ability of the Logistic Model to fit the data.



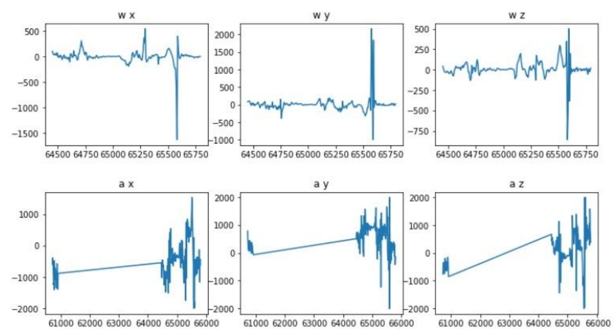
Wearable Device
Outside of Hand Position

The first usage of the device gave some interesting results. An initial connection issue raised its ugly head in the BLE connection. The device frequently disconnected and shut off during data collection. We responded to these issues by rewriting the Python program to automatically reconnect after disconnection. Some connection issues still existed with the firmware and we should investigate them in future implementations of the device. After connecting through the ST library it also seems connecting using the gatttool would allow the device to operate easier. In fact the ST tool appears as a wrapper for gatttool.

As the device sits on the back of the hand and uses BLE, the cradle made a perfect solution for encasing the device. Although difficult to solder the device came right together with a miniature soldering iron. It did however make the device too difficult for the entire team to fabricate. In its current form it provided a sturdy solution for the placement of the device.

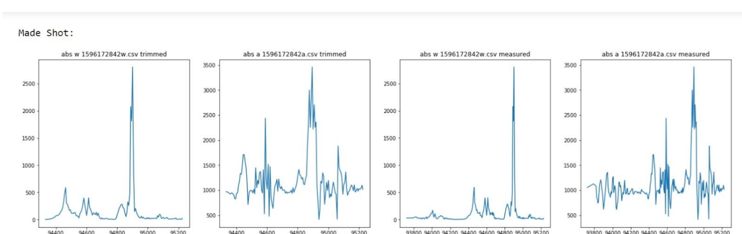
Lessons from Data Acquisition

The first attempt to take data showed the interval used in the sample code of the blue_st_sdk library contained a very small window to take data. It encompassed a couple of milliseconds and did represent some aspects of the data but nothing telling enough to make a prediction of anything beyond numerology. After coming back for a second attempt with a larger window the graphs from the *Finding Classifiers* notebook emerged.



Data Collection Shot Two

Upon looking at the data, several shots contained intervals of missing data. As early as the second shot attempt the device appeared to miss recording some of the data points. As seen in the graphs to the right, with the direct values coming from the device, a long period of missing data occurs in the region 61000-64000. In order to filter the data the conditions on the angular speed and acceleration trimmed the data down to the period during the shot. In the Python notebook *Testing Categorizers* the summary of the data recording shows the results of trimming the graphs. Given this data filtering the results of the shots appear more uniform. In later iterations a low pass filter may make the data more predictable.

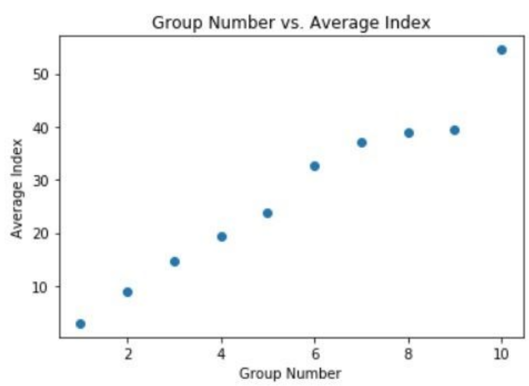


Data Before and After Filtering

After cleaning the data the readily available values didn't give much of a high z-score between the data sets of makes and misses. I chose to use z-score over correlation values due to the Boolean form of the data. The highest z-scores peaked at about 0.25 between populations. Given these results I looked into other features for analysis. As the device always measures the direction of gravity when measuring the acceleration, at the start of data collection you can orient the device to determine the vertical axis. With similar initial hand position this device may determine the three components of velocity by numerically integrating the axis of the device and accelerations through time. Although the hand position should remain the same, alterations of the position will lead to uncertainty in the calculations of the velocity. As a symmetry about the vertical axis exists in this calculation there still remains some uncertainty in the measurement of velocity. Even with the uncertainty in these measurements they still provided quite useful in the differentiating between makes and misses. They contained some of the highest z-scores of the all of the available features

As the motion also has a natural frequency dependent on the distance from the basket (i.e. shooting further away requires a greater speed and a greater frequency with the same shooter) we considered using the FFT to find maximum frequencies of the motion. This yielded some additional features to use in the model.

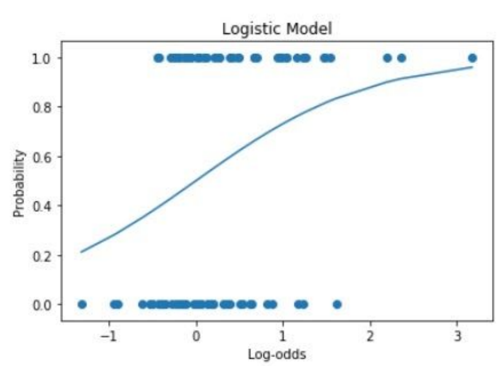
Results of Data Collection



Results of KMeans

As dictated by the physics of the problem the horizontal and vertical components of velocity correctly grouped the shots. As the graph to the left shows KMeans could group the data taken shooting and moving backwards a foot after 5 shots. We let KMeans group the shots and then found the mean index of the groups in the list of shots. KMeans gave almost a completely straight line. With further analysis the results of the type of miss (left, right, short long) included with a more accurate measurement of the components of velocity would give a better grouping.

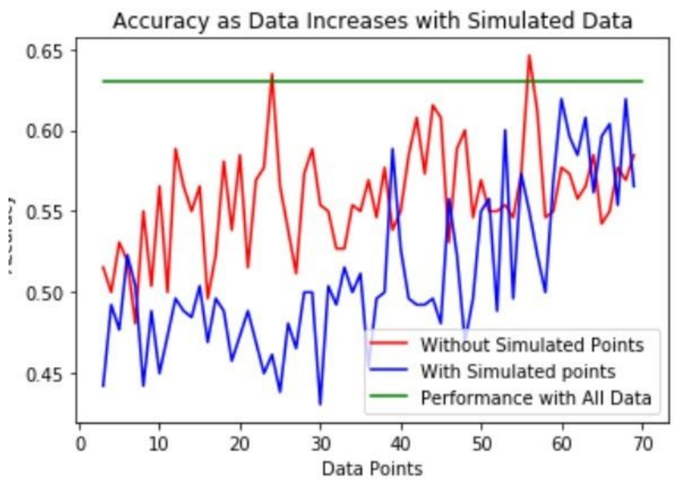
Although these features could determine the distance from a hoop the Logistic Model gave room for improvement with additional models. With the 5 best features to limit dimensionality



Logistic Model Fit

we arrived at a 63% accuracy. By adjusting hyperparameters we achieved a similar accuracy on the data set consisting of the limited measurement time. As these hyper parameters did not replicate that accuracy on the data set consisting of a longer measurement window we suspect it might overfit the data.

Given this data we could now test the effectiveness of simulated data. The chosen logistic model needs at least 5 points with 2 different values recorded for the result of the shot. In the table to the right using a normal distribution for the parameters we simulated data and labeled values outside of one standard deviation a miss. The model appears to outperform the model using only data when about 10%-15% of the data is simulated. The simulated data does not appear to have too much value otherwise. Given the results of this experiment our devicel should only simulate data when there is not enough data for the Logistic Model to converge.



Effectivness of Simulated Data

Ready, Shoot,
Shot Recorded,
Follow Through

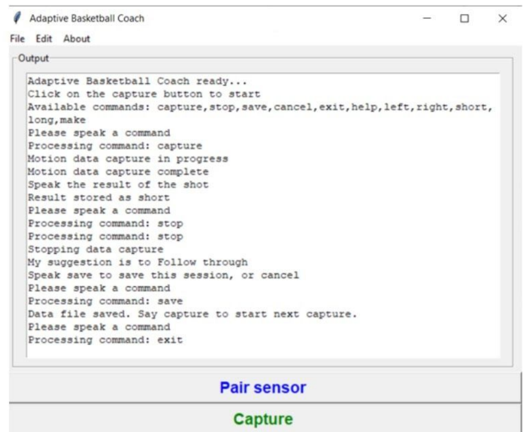
Success!

Our current device does work. Given enough data it will actually make suggestions as you shoot.

Model Refinement

In the initial design we wanted a device you could wear that would give individualized suggestions to improve your shooting. Given the performance of the device we created something that accomplished our goals without making excessive compromises. Even after successfully completing the project, upon finishing the device we felt the project opened more doors than it closed. Opportunities for improving performance and the user experience presented themselves throughout the creation of the device.

The first obvious improvement comes from one of the early compromises we made in the project. When we wrote the code to connect through BLE we noticed the audio connection used a different part of the blue_st_sdk library. This kept audio commands out of this version of the code. If we could rewrite the code we would use a wireless headphone microphone combo working through the Bluetooth connection of a Raspberry Pi or mobile device. In the threads iteration of this device a GUI and Hands Free version exist for future development.



GUI and Hands Free Interface

The second glaring improvement comes from absence of threads. The code originally ran with threads but we removed them early in development to create a running prototype to take data. Part of the team continued to develop using the threads architecture and the other created a synchronous code to start getting data. Right now these code bases do not meet. The GUI also needs to take into consideration the devices that may connect to the Sensor Tile and how they may connect to the Sensor Tile. Given these considerations the device will appear to run faster to the end user.

In addition to design improvements the Logistic Model needs further research. This may be by improving the performance of the model or by considering more features to examine. For instance one of the stumbling blocks for the Logistic Model comes symmetry about the vertical axis. To remove this symmetry the device may take a compass heading. As different team members spent a while working on this data it seems the acquisition of additional data will not improve the model as opposed to finding new features.

Distribution of Duties

Sishir Yeety

- **Data Analytics**
- Machine Learning
- Hyperparameter Tuning
- Logistic Regression
- K Means Clustering

Diana Liu

- **User Interface (UI)**
- Main Program Integration
- Text to Audio Programming
- Architecture Mapping
- Main Program Debugging
- Presentation Slides
- Writeup

Gurleen Bath

- **Data Collection**
- Analyzing Data
- Cleaning Data
- Presentation Slides
- Writeup

Brendan Keegan

- **Data Collection**
- BLE programming
- Analyzing data
- Video 1
- Video 2

Ben Trey

- **Sensortile Programming**
- Data Collection
- Data cleaning
- Modeling
- Data Analysis
- Logistic Regression
- Exploratory Data Analytics

Conclusions

The mechanics of shooting a basketball have been optimized and are well understood. A coach can tell you the motion and position of every body part throughout the shot. A problem arises when taller players or players with limited mobility try to replicate this motion. Most players will adapt their motion to accommodate their body. For instance, Kobe Bryant didn't point his feet at the hoop to allow for the range of motion of his elbow, Shaq can't bend his wrist due to a childhood injury, Durant can't place the ball at a point on his forehead while maintaining a reasonable bend of his elbow, and even Steph Curry places the ball below his chin as he starts his motion. Historically great shooters such as Peja Stojakovic, and Larry Bird incorporated the off-hand, and the only suggestion a coach could give them is to stop using their off-hand. With this in mind, a coach should assist a shooter by understanding their muscle control, stature, and quirks. In this paper, we share the results of our experimental project on a wearable device that combines machine learning and digital sensor processing to provide optimized and customized coaching to the basketball shooter. This paper reports on the vision of our project-Adaptive Basketball Coach (ABC)—by Treaty Bio, its accomplishments, and also some areas for further future improvements.

[1] https://www.basketballnetwork.net/ted-st-martin-holder-of-the-guinness-world-record-for-most-consecutive-free-throws/
[2] https://en.wikipedia.org/wiki/Tom_Amberry
[3] https://www.nytimes.com/2006/05/05/sports/basketball/05levitt.html
[4] https://www.youtube.com/watch?v=4Xh4Y-XYI_4#:~:text=Al%20Callejas%20(Basketball%20shooting%2Dskills,lessons%2C%20clinics%2Cs%20or%20lectures.
[5] https://www.wnba.com/news/elena-delle-donne-consecutive-free-throws-made-streak-wnba-record/#:~:text=The%20NBA%20record%20for%20consecutive,Abdul%2DRauf%20(81).
[6] https://sportsnaut.com/2020/03/watch-sixers-shake-milton-breaks-nba-record-for-most-consecutive-threes/#:~:text=It%20starts%20with%20Stephen%20Curry,14%20consecutive%20made%20three%2Dpointers.
[7] FOWLES, G. R. (1999). *Analytical mechanics 4th ed*. Philadelphia, Saunders College Pub.
[8] https://github.com/STMicroelectronics/BlueSTSDK_Python
[9] https://sites.google.com/view/ucla-stmicroelectronics-iot/home
[10] https://www.st.com/en/evaluation-tools/steval-stlkt01v1.html