March 2017

# Passive Biometric Authentication via Head Mounted Display using Ballistocardiography

Jesse Kyle Earisman
*Worcester Polytechnic Institute*

Joshua Aidan Hebert
*Worcester Polytechnic Institute*

Sai Kiran Vadlamudi
*Worcester Polytechnic Institute*

# Passive Biometric Authentication via Head Mounted Display using Ballistocardiography

---


Submitted To:
*Project Advisor:* Krishna Kumar Venkatasubramanian, WPI Professor

*Submitted By:*
Jesse Earisman
Joshua Hebert
Sai Kiran Vadlamudi

Date: March 20, 2017

**Abstract**

*We demonstrate that by monitoring readings from the accelerometer and gyroscope of a head-mounted display, we are able to construct a waveform that is closely tied with the cardiac cycle of the wearer. Furthermore, we show that, from this waveform, we can then extract features that are not only consistent over time in the wearer, but also reasonably unique between different individuals. By then constructing an ensemble of random forest classifiers, we show that such a model can be used to determine if a new set of features does or does not belong to wearer. In this way, such a system can be used in an authentication context with a high degree of accuracy.*

# Contents


I Introduction . . . 3
II Background . . . 4
III Problem Statement . . . 6
IV Approach . . . 6
V System Model . . . 7
- i Collection Device . . . 7
- ii Data Collection Software . . . 8
- iii Collection Process . . . 9
- iv Threat Model . . . 9

VI Parameter Tuning . . . 10
- i Baseline . . . 10
- ii Preliminary Evaluation Methods . . . 11
- iii Feature Extraction and Selection . . . 13
- iv Window Size . . . 15
- v Machine Learning Model Tuning . . . 16
- vi Random Forest Parameter Tuning . . . 17
- vii Training an Ensemble Classifier . . . 19

VII Experimentation . . . 20
VIII Metrics . . . 22
IX Results . . . 23
- i Final Parameters . . . 23
- ii Evaluation Methods . . . 23
- iii Final Results . . . 24
- iv Explanation of results . . . 25

X Discussion . . . 25
XI Future Work . . . 26
XII Related Work . . . 26
- i BioGlass . . . 26
- ii BioInsights . . . 27
- iii Existing HMD authentication methods . . . 27

XIII Conclusion . . . 28

**Appendix A Google Glass** **31**

**Appendix B IRB Form** **32**

# List of Figures


1 The mechanism used by Issac Starr for generating a Starr or High Frequency BCG [13] 4
2 Starr and Nickerson BCG . . . . . 5
3 Ultra-low and Direct Body BCG . . . . . 5
4 Overview of System . . . . . 7
5 Sequential System Model . . . . . 8
6 Data Transfer Protocol Overview . . . . . 9
7 Setup for Data Collection during Study . . . . . 9
8 Translation planes for HMD accelerometer . . . . . 10
9 Rotational axes for HMD gyroscope . . . . . 10
10 Steps in BCG extraction from gyroscope/accelerometer readings . . . . . 12
11 Example shape feature with BCG points HIJKL highlighted . . . . . 14
12 Results of analysis of window size . . . . . 15
13 Comparison of machine learning models . . . . . 16
14 RandomForest Parameter Tuning . . . . . 18
15 Majority Voting Results with the three overlapping sets of 35 subjects . . . . . 19
16 Ensemble classifier example . . . . . 20
17 Experiment demographics . . . . . 21
18 Split testing (Subject 1 being the only who should be accepted) . . . . . 23
19 Unseen testing (Subject 1 being the only one who should be accepted) . . . . . 24

A.1 Google Glass HMD[2] . . . . . 31


# List of Tables


1 Data Collection Protocol Specification . . . . . 8
2 Feature set vs accuracies (20 Subjects) . . . . . 15
3 Split testing results . . . . . 24
4 Unseen testing results . . . . . 24

# I. Introduction

Computing Technology is the stamp of this era similar to manufacturing in the last era and as such it is ever-present in our society. As with any popular field the involvement of the general populous in the field will bring along the attention of the malicious. Fighting against this has been the focus of cybersecurity experts for almost as long as the field has existed. Whether it be through memory, innate feature or genetic characteristics, there has been much research performed to find the balance between security and convenience. After all technology and machines are there to make the lives of humans more convenient and adding cumbersome security practices will only lead to frustration.

Acquiring a balance between security and convenience led to the popularization of passwords which are meant to be something only known to the individual of the account to prevent malicious attackers. This is an ideal solution if humans are as capable as machines in regards to storing and recounting arbitrary strings but that is not the case. Passwords are created by humans who are predictable and emotional thus leading to weak and almost useless security. While this is still the most common security practice, biometric authentication is eroding the share of passwords .

Biometric authentication or using the genetic variances among individuals as an identification method is a widely active field as it relies on almost no mental capacity and is convenient for the users. The decreasing price of fingerprint technology along with the increasing price of devices led to the scanners being on almost all devices. Regardless of the price this is still an additional sensor that needs to be purchased and integrated into devices. This sensor requirement means it cannot also be incorporated into some of the wearable devices like smart watches. Additionally the technology is problematic in wearable devices as it would require user interaction in potentially unnaturally manner, touching a watch, touching a glass on the side of the head etc.

This work demonstrates a way to incorporate the convenience of biometric authentication using the heart as a data source rather than the fingers. The impact of this is that authentication can be performed anywhere on the body and required no user interaction with the device. Our focus especially is the increasing head mounted display market, exploratory devices, such as Google's Glass or virtual reality devices like Oculus Rift and HTC Vive. All of these leverage similar technology to present information to the wearer and as such, they also suffer from the same difficulty in authenticating the user. The positioning of the devices on the user's face poses difficulty for conventional authentication methods, such as fingerprint or password.

We propose an alternative method of authentication based upon the readings of cheaper and already required sensors, accelerometer and gyroscope. Along with this data we show that uniquely identifying cardiac information can be extracted. Leveraging machine learning we train an authentication system to recognize whether or not the wearer of the device is the owner. Furthermore we demonstrate that this process can, with a sufficiently trained model, classify a wearer with only a few seconds of collected gyroscope/accelerometer data making it a realistic system.

We use cross validation to ensure that our results are generalizable and also test with data from subjects who are never introduced to the system. We show the applicability of this system due to low false accept and false reject rates. Our testing also shows this system holding even with a statistically large pool of users, and within such a group it still remains possible to verify the identity of a user with a degree of accuracy not reached in previous studies in the field.

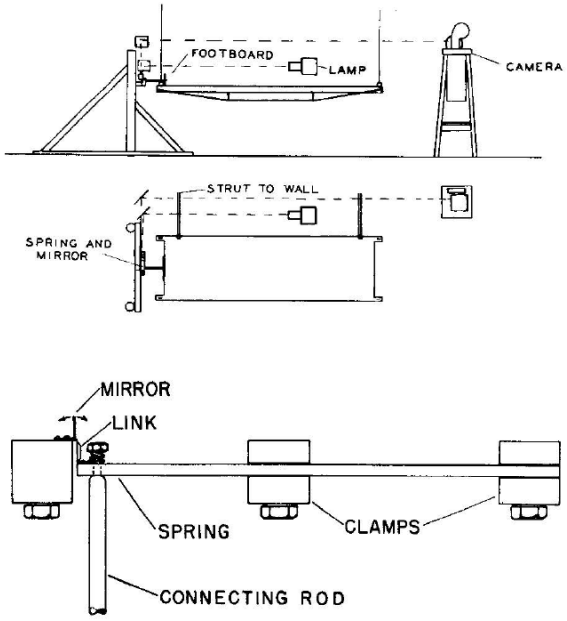


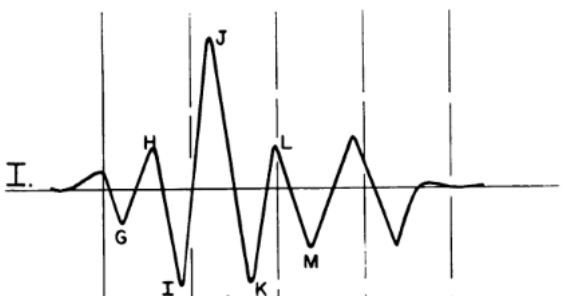

**Figure 1:** *The mechanism used by Issac Starr for generating a Starr or High Frequency BCG [13]*

## II. Background

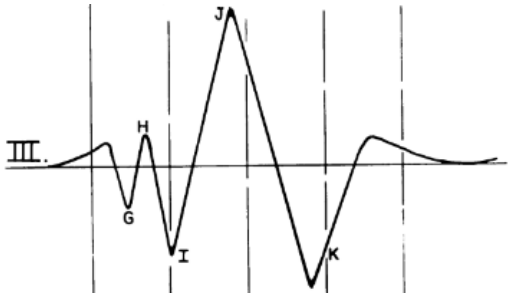
Before discussing our system it is important to discuss the theory behind it, Ballistocardiography, along with the basic explanation of the two phases of the cardiac cycle. The Diastolic phase, when the heart is resting and the chambers are simply being filled with blood and the Systolic or ejection phase, when the heart contracts to pump the blood throughout the body. The force needed to pump the blood travels throughout the body resulting in a subtle but observable movement. This movement is the foundation for the field of Ballistocardiography which started in 1877 but mainly boomed in the mid 20th century [11]. During the early stages of the field there was not an established method for collecting this movement data so different techniques were developed by the various researchers. Slight variations in the measuring techniques resulted in differences in output waveforms or Ballistocardiograms. The four recognized and discussed are high-frequency, low-frequency, ultra-low frequency and direct-body [12] Ballistocardiograms, hereafter referred to as BCGs.

The high frequency or Starr BCG is recorded with the help of the apparatus in **Figure 1**. The output produced from this machine, see **Figure 2a**, is the illustration typically associated with a BCG due to the seven distinct peaks. The seven peaks occur with different phases in the cardiac cycle and thus are important in analyzing the waveform. The G, H, I, J, K peaks occur during the Systolic phase and the L, M and N peaks occur during the Diastolic phase. One of the more important of these peaks is H, which "begins its ascent...near the peak of the [ECG] R wave"[12]. This allows identification of the start the beat cycle in a BCG without the need for a corresponding ECG. Another interesting fact is there are only standardizations and recommendations for the displacement and not velocity or acceleration data.

The low frequency BCG, commonly referred to as a Nickerson BCG, also uses a suspended bed design like the Starr machine. The output graph is similar to that of the Starr BCG and contains

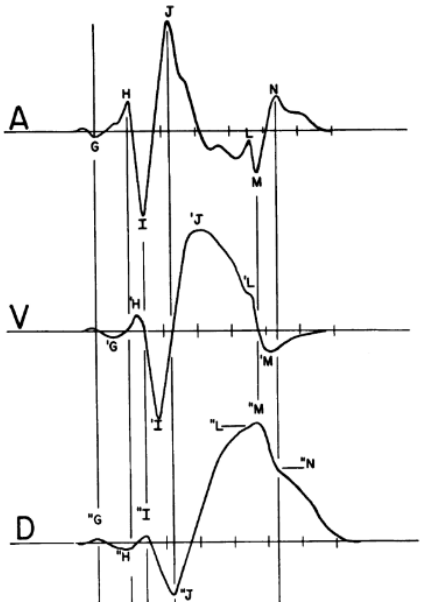



**(a)** *A sample High Frequency BCG, Starr BCG [12]*

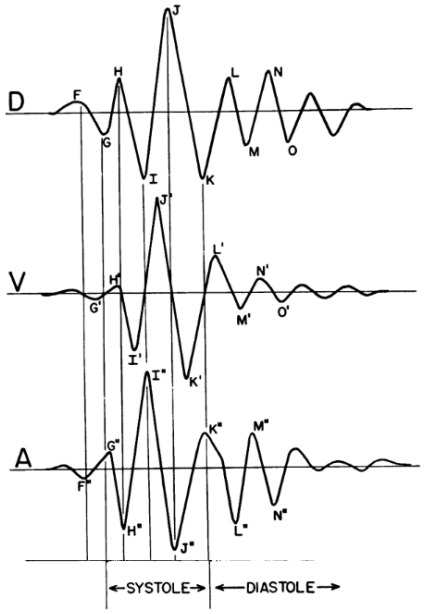



**(b)** *A sample Low Frequency BCG, Nickerson BCG [12]*

**Figure 2:** *Starr and Nickerson BCG*




**(a)** *A sample Ultra-low Frequency BCG with acceleration, velocity and displacement [12]*




**(b)** *A sample Direct Body BCG with acceleration, velocity and displacement variants [12]*

**Figure 3:** *Ultra-low and Direct Body BCG*

the same Systolic peaks but do not contain the L and M Diastolic peaks, see **Figure 2b**. Another difference is that the timing between these peaks is different from a Starr BCG in that some of the peaks, J and K, occur "considerably later in time" [12]. The J occurs at end of systole/early diastole instead of mid systole and the K wave occurs in mid diastole instead of end of systole [12]. Also unlike in a Starr BCG, the baseline of a Nickerson BCG is affected by Respiration of the subject [12]. Similar to Starr BCG only displacement data seems to be standardized and recommended.

The Ultra-low frequency BCG uses a suspended platform design similar to Starr's. These types of BCGs can be observed as acceleration, velocity and displacement varieties, see **Figure 3a** and the differences between each are described. The acceleration variant is different in that the K peak may be absent or extremely subtle and the other peaks occur slightly earlier in time, also there may be additional peaks [12]. Similar to a Starr BCG, respiration does not affect the baseline in

the acceleration variant. The velocity variant contains three distinct peaks I, J and M where the J wave extends to or slightly beyond the end of systole and the M peak is in early diastole [12]. This variant is also affected by respiration and the peaks occur later than in the displacement and acceleration BCGs [12]. The displacement variant is similar to the acceleration variant in regard to the times of the peaks but there are only two apparent peaks and respiration affects the baseline.

The direct body BCG, see **Figure 3b**, is different from the others in that there is no large machine where the subject lies which records the movement of the whole body. Instead the subject simply lies on a flat surface, like the floor, and puts on a single instrument which measures the motion of a single part of the body [12]. An example given is "a crosspiece connecting the two shins" [12]. Standard identifiable properties are given for acceleration, velocity and displacement variants of BCGs this type like those from the UF-BCG. In the displacement variant the distinctness of the peaks is similar to a Starr BCG but unlike a Starr BCG the peaks themselves appear later in time and the baseline is affected by respiration [12]. The velocity variant is similar to the displacement variant in identifiable peaks but some of the peaks appear earlier in time and the magnitude of some of the peaks is smaller than the corresponding ones in displacement variants [12]. The acceleration variant is similar to the displacement variants in the peak timings but the direction of them is reversed and respiration has no effect on the baseline. The waves themselves are less smooth and sharper [12].

The differences between these four main types of BCGs seems to be coming from the techniques used in acquiring the data which therefore results in slightly varying types of outputs. The main interest in any type of BCG seems to be coming from a set of peaks and the stages they occur in the systole-diastole cycle. The Scarborough paper seems to be an attempt to provide a standard set of observable features for all of the types as to make the field more consistent. Our technique for generating a BCG is similar to the Dock-BCG in that the motion of a single body part, the head, is recorded and doesn't require the user to lie in an apparatus.

## III. Problem Statement

We are attempting to solve the problem of authentication on head mounted displays, and possibly other wearable technologies. Due to the nature of these devices, traditional authentication mechanisms such as passwords can be cumbersome or impossible. Our goal is to develop a system that requires minimal interaction from the user, while simultaneously achieving a high degree of accuracy. Additionally, it is desirable for our system to be able to continuously and passively authenticate the user while they are using the device.

## IV. Approach

In this section, we introduce a process for authenticating an individual wearing an HMD based on characteristics of their BCG. By monitoring subtle tremors in the head of a wearer via onboard accelerometer and gyroscope, our system can construct such a waveform; we then leverage the uniqueness of this waveform between individual users to classify the wearer. In this way, our system is able to authenticate users. In order to construct a fully fledged authentication system we divide this system into two stages, the training stage and the authentication stage (see **Figure ??**). In the beginning of each, we read the accelerometer/gyroscope readings of the HMD to derive a BCG from which unique features are extracted. The process involved during each of these stages is described below.

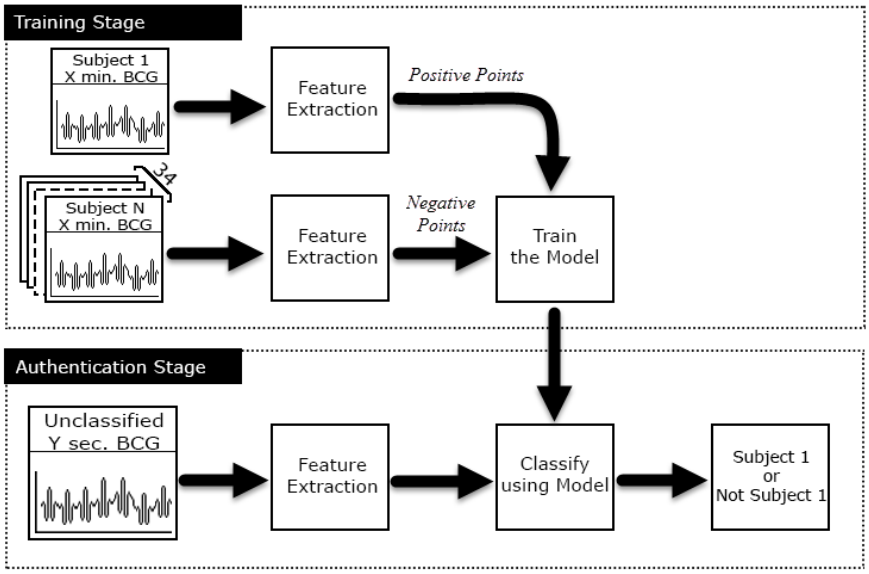



**Figure 4:** *Overview of System*

**Training Stage:** The system first collects a sufficient number of data points from the user to reasonably ensure that they represent the user's normal state, and that anomalous points are scarce enough to not cause issue. For the purposes of this experiment, we chose this period to be 10 minutes, as this mirrors similar work done in the field[8]. The BCG and subsequently, identifying features, are then constructed from this. Using these features, we then train a machine learning model to classify the collected data points as belonging to the wearer. Simultaneously, we mix in data points collected from other users to provide the complement to this set. At the end of this process, we are left with a model that is able to authenticate a user. We repeat this process to construct multiple models for the same person, using different samples from other users for each. At the end, we are left with an *e*nsemble of classifiers, in which each is capable of classifying the user.

**Authentication Stage:** Utilizing the models created during the training stage, the system is then able to authenticate a user. By sampling a small amount of data from the user, this new point is presented to each model in the ensemble. If enough models classify affirm that the wearer is the owner to meet some threshold, the system will accept the wearer and authenticate them.

## V. System Model

Here we describe the system we used to collect the data from the user and transmit it to the machine where authentication computation is done.

### i. Collection Device

We utilize an HMD that provides a three-axis accelerometer and three-axis gyroscope, each sampling at a minimum rate of 50Hz. Between these two sensors, we were able to read six discrete data streams: the three axes of the accelerometer and three axes of the gyroscope. For reference,

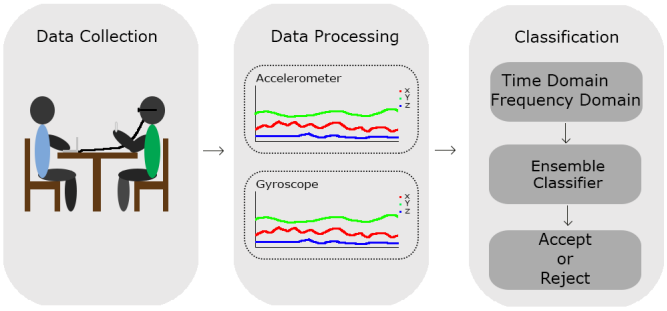



**Figure 5:** *Sequential System Model*

the orientation of the accelerometer is shown in **Figure 8**, and that of the gyroscope is shown in **Figure 9**.

## ii. Data Collection Software

There exist two components in the data collection setup: the HMD itself and a nearby laptop, which is used for storing and processing the data. These two devices were on the same local area network, connected via 2.4GHz 802.11g wi-fi. The software architecture of these components is described below:

**Laptop Software:** In this model, the laptop acts as a server, listening for reports from the remote HMD client. This is achieved via a simple HTTP server written in NodeJS[4] listening for POST requests. each request is simply a JSON array of records, where a record is described in the **Table 1**. The server then writes each record as a line in a CSV file on disk, where it can then be freely read.

| Field Name | Description | Data Type | Possible Values |
|---|---|---|---|
| TYPE | Name of sensor | String | GYRO, ACCEL |
| TS | The timestamp in microseconds | Long Integer | From 0 to $2^{63} - 1$ |
| X | Movement in the X component | Float | From $2^{-126}$ to $(2 - 2^{-23}) * 21^{27}$ |
| Y | Movement in the Y component | Float | From $2^{-126}$ to $(2 - 2^{-23}) * 21^{27}$ |
| Z | Movement in the Z component | Float | From $2^{-126}$ to $(2 - 2^{-23}) * 21^{27}$ |

**Table 1:** *Data Collection Protocol Specification*

**HMD Software:** We developed a simple application that read data from the onboard accelerometer and gyroscope as rapidly as possible, formatted it into the protocol described in **Table 1**, and sent it as an HTTP POST request to the laptop. It is important to note that the timestamp field is populated by the HMD; as our two components communicate via TCP, each record is guaranteed to be delivered, meaning that network latency does not affect any numbers. These two operations (reading data and sending data), are carried out asynchronously on two separate threads, with inter-thread communication buffered internally such that network difficulties do not cause sensor data reading to stall nor cause lost samples.

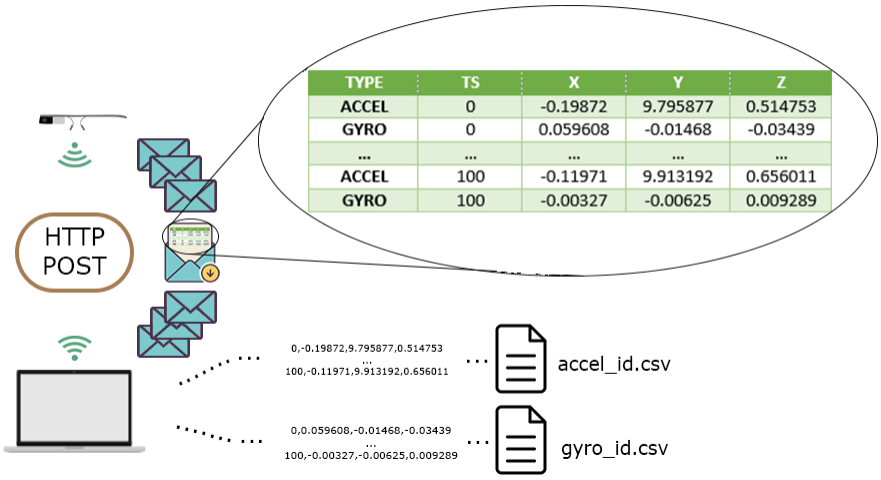



**Figure 6:** *Data Transfer Protocol Overview*

## iii. Collection Process

The setup for our study can be seen in Figure 7. We gave the participants the HMD to wear, and they were asked to assume a seated position and hold still for a period of 10 minutes. During this time, the HMD relayed readings from the accelerometer/gyroscope wirelessly to a nearby laptop, where the readings were stored. To alleviate boredom and limit any fidgeting that could have skewed results, we allowed the subjects to browse their mobile devices during the collection period. We forbade them from talking or otherwise moving their heads in any significant manner.

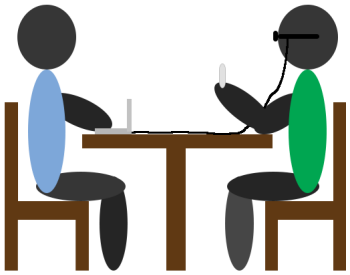


**Figure 7:** *Setup for Data Collection during Study*

## iv. Threat Model

The threat to our authentication system is a malicious user attempting to gain access to a secure head mounted display. By placing it on their head and attempting to authenticate as the normal user. Our system requires a few reasonable assumptions about the capabilities of the attacker.

1. The attacker has not modified the authentication software running on the device

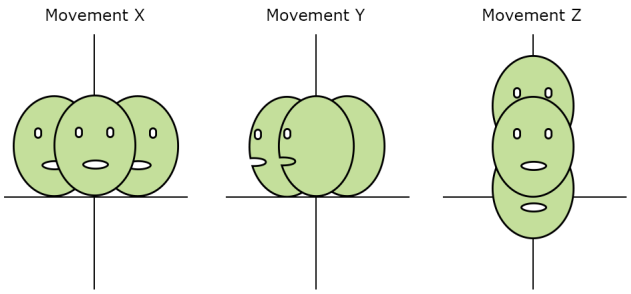



**Figure 8:** *Translation planes for HMD accelerometer*

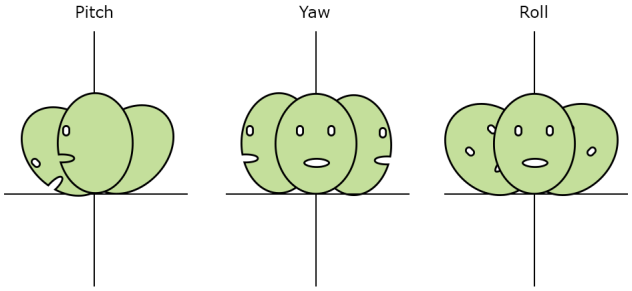



**Figure 9:** *Rotational axes for HMD gyroscope*

2. The attacker does not have access to the normal users biometric data.
3. The attacker cannot modify the sensor readings in any way during the authentication process.

## VI. Parameter Tuning

In determining the specifics of our approach, we empirically explored different possible values for each parameter to select the ideal one for each. The process is described here below:

### i. Baseline

Our goal is to derive a ballistocardiogram (BCG) from the accelerometer and gyroscope readings and use it to build an authentication system. There is a thriving research community in this area and our baseline consists of the results from Hernandez et al.[8] study. This study along with the BioInsights[8] provided us initial insight into the known best parameters for computing a BCG along with the initial feature set and analysis techniques. The initial processing steps are as mentioned below and are identical to the steps followed in the BioInsights[8] study.

**Alignment and Interpolation** In order to make the data collected usable in calculation, it first needed to be preprocessed. This is due to a number of reasons. Primarily, sensor data from any Android device is not guaranteed to align exactly to a particular sample rate. A sample rate of

50hz could yield intersample gaps of anywhere from 5 to 20ms. Additionally, there exists no guarantee that the samples recorded from the gyroscope and accelerometer are synchronized or aligned in any way. In order to address these two concerns, we modified the data in the following way.
Firstly, we truncated the beginnings and endings of both the gyroscope and accelerometer such that the first and last samples of each were as close as possible to their corresponding sample in the other sensor. We then utilized SciPy's [9] interp1d function to interpolate the data to 100Hz and align samples with one another (e.g. the first samples of the gyroscope and the accelerometer then shared the same timestamp, and so on with all subsequent data points). We were then able to utilize this data in conjunction with a process similar to that used in the BioInsights[8] to construct a BCG.

**Segmentation:** The next step was to divide the preprocessed input into discrete segments, which we could then extract features from. We chose a window size $w$ and segmented the data into blocks of $w$ seconds. Our initial window size was chosen to be 10 seconds, meaning each 10 minute sample generated 60 segments. All further processing was done on a per-segment level. This guaranteed that features extracted from one segment remained independent of other segments from the same sample.

**Obtaining Ballistocardiograph Waveform:** An example of a raw reading of the accelerometer channel can be seen in **Figure 10a**.Each of the sensor components (6 total) in the segment was normalized to have 1 mean and unit variance, as shown in **Figure 10b**. A rolling average filter of 35 samples was subtracted from each component, to correct for large motions as well as gyroscope and accelerometer drift, as shown in **Figure 10c**. A butterworth band-pass filter of order 4 with cutoff frequencies of 4-11Hz was applied to each sensor component. The results can be seen in **Figure 10d**.

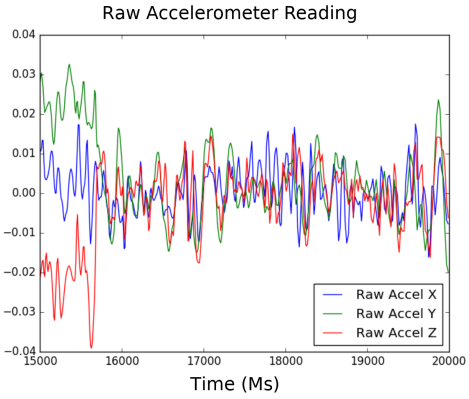
**Sensor Selection:** After processing, we had obtained 3 dimensions of gyroscope data and 3 dimensions of accelerometer data, for a total of 6 sensor components. In order to progress, we needed to reduce our set to a single waveform. Following the precent set by [8], we applied a fast fourier transform to each component, and the final BCG waveform was selected as the component with the highest amplitude response in the frequency domain, as seen in **Figure 10e**. The graph of the frequency domain on the same component was also saved, such that we may extract features from it later.

## ii. Preliminary Evaluation Methods

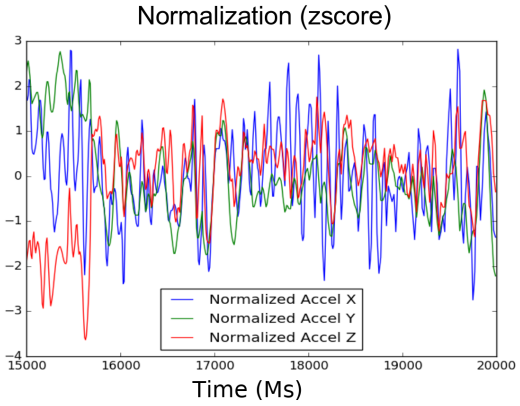
After implementing the functionality necessary to compute ballistocardiography waveforms from accelerometer and gyroscope data, next step is to test the system to make sure the baseline results are being achieved. Before continuing on to the feature selection process, it is important to discuss some of the tools and terminology to better understand the approach.

### Weka

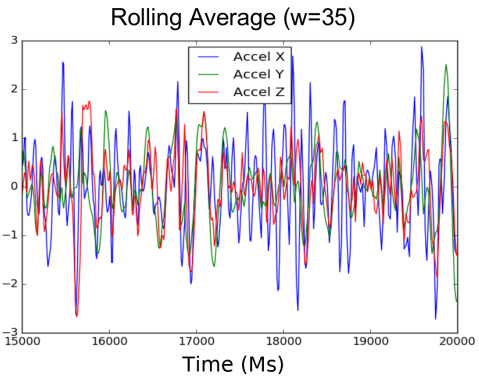
Weka[6] is piece of software that implements several machine learning algorithms. It allows users to input custom data and see the results of training and testing custom classifiers based on the



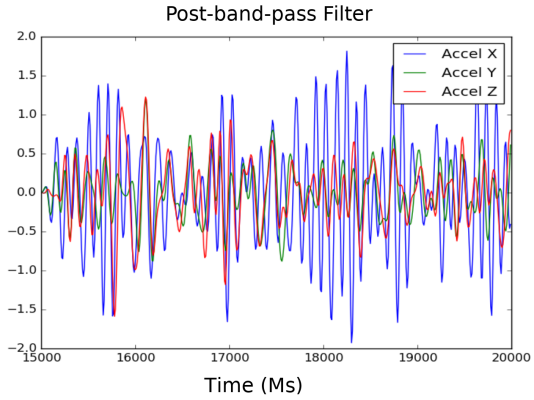

**(a)** *Raw accelerometer reading*



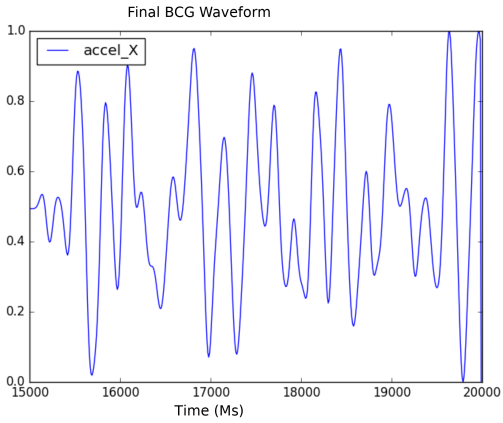

**(b)** *Normalized accelerometer reading*




**(c)** *Post-rolling average filter lines*




**(d)** *Accelerometer signal after applying butterworth band-pass filter (4-11Hz)*




**(e)** *Final BCG after FFT component selection*

**Figure 10:** *Steps in BCG extraction from gyroscope/accelerometer readings*

data. Using a machine learning tool designed to speed up the trail process, Weka, we ran our system with the following setup to get the balanced accuracy.

**Random Forests (RF)**

Random Forests are a commonly used machine learning model[3]. Given a selection of positive and negative feature points, they randomly create a number of decision trees that each use some subset of the features available, ordered randomly. New features are then fed into this "forest" and classified by majority rule. For preliminary evaluation, we used the default random forest classifier in Weka. For each subject, we created a single classifier. The first 8 minutes of that subject's data were used as positive training data. The negative training data was selected by choosing random points from other subjects. The model was trained with 2 negative feature points for each positive point.

**Cross Validation**

Cross validation is a technique for estimating the accuracy of a classifier. In each round of cross validation, input data is divided into K discrete subsets. K-1 subsets are used to train, and the other is used to test the resultant model. For assessing the accuracy of our system, we used 10-fold cross validation, which divides the data into 10 sets, and performs 10 rounds of cross validation. Each round uses a different single subset of data as the testing set. Weka has the ability to perform cross-validation automatically.

**Balanced Accuracy**

Balanced accuracy is defined as the ratio of correct classifications over all classifications. It served us as a quick way to assess our system's performance in lieu of a more in-depth analysis. Weka is capable of automatically calculating Balanced Accuracy. We used the balanced accuracy as a guide during our preliminary analysis to help identify the optimal setup.

## iii. Feature Extraction and Selection

Once the BCG waveform was generated, we began searching for features to extract for our machine learning model. The first features we extracted were inspired by the work done in BioInsights[8]. First, we identified points H, I, J, K, and L in the waveform, these points correspond to the points in **figure 11**. We located these points by first segmenting the entirety of the collected data into windows of size $w$. Within each window, we found the peak with the largest positive amplitude and treated it as the central peak J. We then located next and previous valley and peak, yielding 5 points. As a result of this technique, for each window, our methodology identifies a single cycle, regardless of how many cycles total exist within the window. We believe that using the most prominent cycle yields features with the most information. For each cycle, we calculated the distance and angle between each pair of points, for a total of 10 distance features and 10 angle features.

Using these "shape" features– as they describe the shape of the waveforms, we performed initial testing to assess viability. With 20 participants, a 4 second window, and using a random forest classifier, tenfold cross-validation yielded an average balanced accuracy of 88.54%. While the results here were promising, we sought to explore additional avenues.

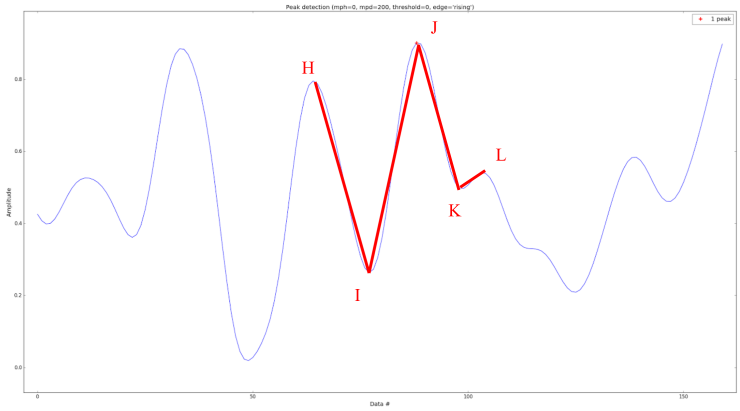



**Figure 11:** *Example shape feature with BCG points HIJKL highlighted*

Having already calculated the fast Fourier transform (FFT) of the BCG, we began to search for features in the frequency domain. Two features we tested first were entropy and energy. We decided to use these features because they were simple to calculate, and we believed they would succinctly summarize the features of the BCG. The formula for these is shown below.

Given the FFT of each window, we calculated the energy and entropy and added these two features to our feature list, in addition to the shape features.

Let N be the number of data points in the fast fourier transform F of the BCG, and let $m_j$ be an element of F; then energy $E$ can be described as:

$$E = \sum_{j=1}^{N} (m_j^2)$$

.

In the same domain, let $P$ be a histogram of $n$ bins describing $F$, and let $P_j$ be the value of bin $j$ in $H$; then entropy $K$ can be described as:

$$K = -\sum_{j=1}^{n} (P_j * \log(P_j))$$

In addition to frequency domain features, we also sought to extract more information from the time domain. We looked at several BCG waveforms and believed that the number and amplitude of peaks in the waveform could yield valuable information. We calculated the following features within the time domain of the BCG to see if there was a significant improvement.

**Peak to Peak (PtP) Distance**: The peaks are defined to be any point that has a slope of 0, local minima and local maxima both count as peaks for our purposes as otherwise we would be wasting data. Given this measure, we collected all interpeak distances for consecutive peaks.

**Peak Amplitude**: This measure is the height of a point that has a slope of 0, both local minima and local maxima, from a normalized baseline that will be the center of the wave in

terms of height. Similar to PtP Distance, we collected all peak amplitude measures within a given window.

**Peak Count**: This is the simple count of the number of points that have a slope of 0 with both local minima and maxima within the window.

Using $E$, $K$, shape features, and time-domain features as features, repeating the same process as with the initial testing of just angle and distance measures, we found the average balanced accuracy to increase to 98.62%. Given this, we postulated that shape features were having a negative effect on our accuracy. As such, we repeated the experiment with shape features omitted, yielding a 99.18% balanced accuracy rate with the same data (see **Table 2**). Further testing of feature removal showed that our accuracy consistently improved until a final set of five features was reached. These features were: energy, average PtP distance, average peak height, standard deviation of peak height, and number of peaks. These features became the final set used for the full experiment.

| Features | False Accept (%) | False Reject (%) | Balanced Accuracy (%) |
|---|---|---|---|
| Frequency Domain and Time Domain | 1.14 | 0.5 | 99.18 |
| Energy and Time Domain | 1.02 | 0.5 | 99.24 |
| Shape Features | 15.17 | 7.75 | 88.54 |

**Table 2:** *Feature set vs accuracies (20 Subjects)*

## iv. Window Size


**(a)** *False Accept and Reject by Window Size*

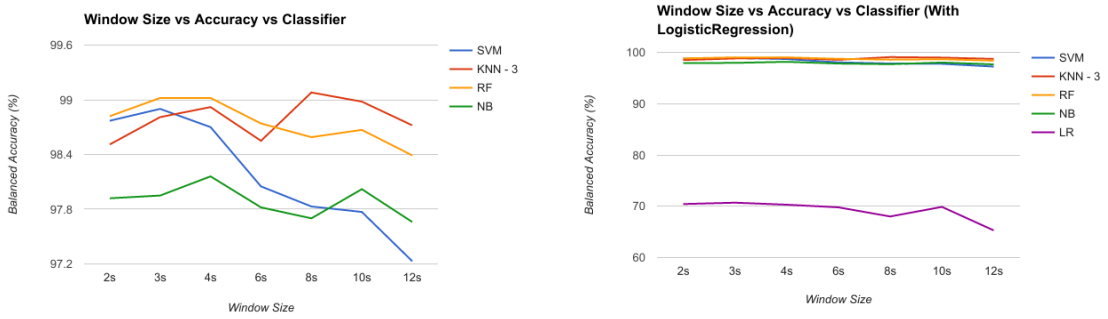
**(b)** *Relationship between window size and accuracy*

**Figure 12:** *Results of analysis of window size*

Through feature selection, our focus was also brought to the concept of the window size and it seemed to be a influential parameter of this model and as such we wanted to test for any improvements. Using the five features that we found to be the best, we investigated window sizes from 2 to 12 seconds in 2 second intervals. We found 2 seconds to be the minimum window size

where we found at least one full heart beat cycle for majority of subjects and going by smaller increments would not give us comprehensive features per beat. The results of this parameter testing can be seen in **Figure 12b**. Again, this balanced accuracy comes as a result of a tenfold cross-validation test of each sample in our sample pool.
We also evaluated the FAR and FRR, and graphed the results in a box plot, seen in **Figure 12a**, by splitting our sample pool. We used the data points of 15 individuals to construct our models, and then used the remaining 5 to test that these unseen 5 would all be (correctly) rejected by the constructed models. This information can be used to validate the cross validated results.
It is evident from this figure that a window size of 4 seconds represents the optimal window size for mitigating the error rate in the classifier. While this is not the fastest authentication time, it is better than the 10-second segments used in the BioInsights study[8]. Maintaining the higher level of accuracy in a faster system is already much better than previous work but there is certainly more tuning that can be done.

## v. Machine Learning Model Tuning




**(a)** *Without LogisticRegression*

**(b)** *With LogisticRegression*

**Figure 13:** *Comparison of machine learning models*

Once the feature extraction process is completed, one file is generated for one set of accelerometer and gyroscope files. Each line in the feature file corresponds to one four second segment of data, containing values for the five selected features. In order to use these features in the context of authentication, we utilized machine learning models, which allow us to computationally categorize data based on arbitrary values.
Naturally, there are many ways to construct such models, and it is inevitable that one type will be better for this specific use case than others. In the previous work the algorithm of choice was the Support Vector Machine but we wanted to try others and search for improvements. Our initial testing through Weka showed RandomForests generally giving the best results but we performed the following testing to verify our choice. Given this, we chose to assess the viability of the following five types of models:

**Support Vector Machine (SVM):** A non-probabilistic binary linear classifier that attempts to divide points provided from training data into two distinct groups, represented by a point in space. Each feature used adds another plane to the space. By placing new points in this space,

the model can then pick which group it thinks the point belongs to.

**K Nearest-Neighbors, where K = 3 (KNN-3):** A lazy learning classifier that defers computation until classification by not attempting to group points at training time. Similar to an SVM, KNN-*X* places all training points in a space. In order to classify a new point, it selects this new point's *X* nearest neighbors and assigns the point their group, chosen by majority.

**Naive Bayes (NB):** Simple probabilistic classifier that operates by applying Baye's Theorem between the features.

**Logistic Regression (LR):** Uses a statistical regression model to assess which category any given point falls into, given the training data.

We trained each model by selecting a set of positive points and negative points from the data, where positive points come from the user the model is accepting, and negative points come randomly from other users, which the model should reject. With an 8 minute training sample and a 4 second window, we were able to generate 120 positive points per user. We chose to match these in a 1:2 ratio with negative points, so from the remaining users' samples, we randomly selected 240 negative points. All five models were created using the Python Library sklearn [10] using default parameters. We repeated the experiment used to select window with each of these models, and the results can be seen in **Figure 13**. Random forests seemed to yeild the highest accuracy.

## vi. Random Forest Parameter Tuning

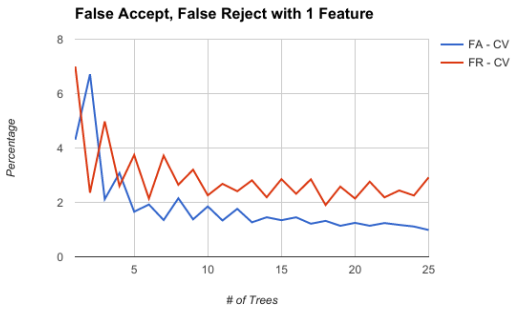
While default parameters are generally the best as they are chosen after thorough testing and complete knowledge about the particular implementation. We didn't want to leave this area untested and so tweaked and observed the results of the two following parameters:

**Number of Trees:** The number of trees that are randomly generated in the forest. Generally the more trees there are the more features combinations will be represented. Due to our small feature set size we wanted to see if altering the number of trees will ensure higher coverage and benefit accuracy.

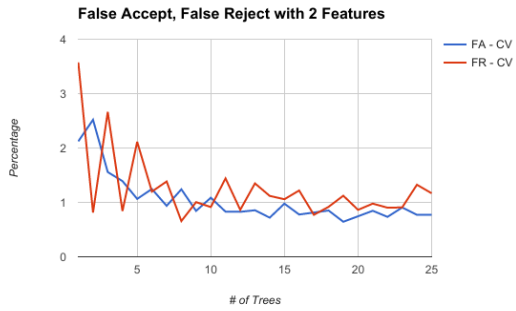
**Number of Features per Tree:** Ensuring coverage can also be done by manipulating this to make sure more features are randomly selected per tree giving a higher coverage with potentially lower number of trees.

We manipulated the number of trees to go from 1 to 25 and the number of features per tree from 1 to 5. Performing cross validation with the models generated as per the above parameter configurations gave us the following results.

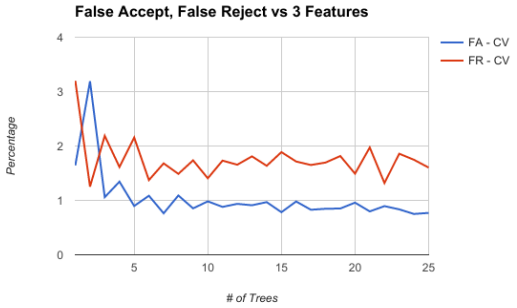
As can be seen with **Figure 14** only trees with 3 features seem to be generating equivalent false accept and false reject rates. The others seem to be diverging error rates instead of converging to an equivalent point. This let us to select 2 features per tree as the optimal. In terms of number of trees 20 trees resulted in a high balanced accuracy of 99.2% while simultaneously giving a low false accept rate of 0.74% resulting in us picking 20 as the optimal number of trees per random forest.



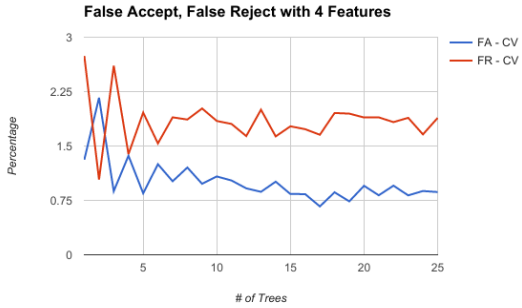

**(a)** *1 Feature per Tree*



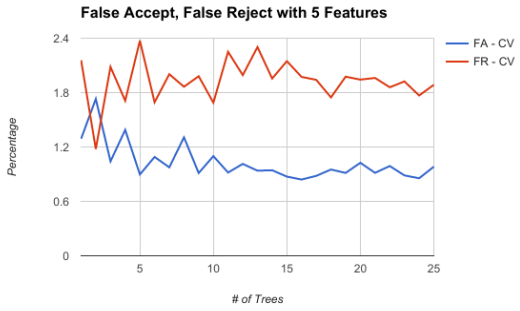

**(b)** *2 Features per Tree*




**(c)** *3 Features per Tree*

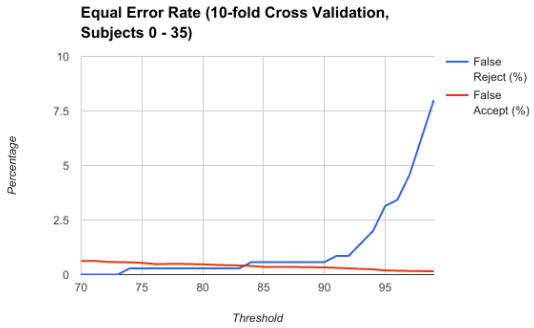



**(d)** *4 Features per Tree*

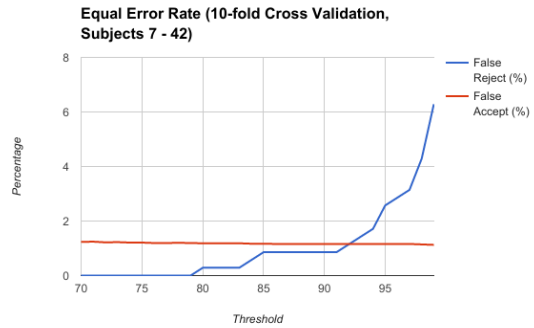



**(e)** *5 Features per Tree*

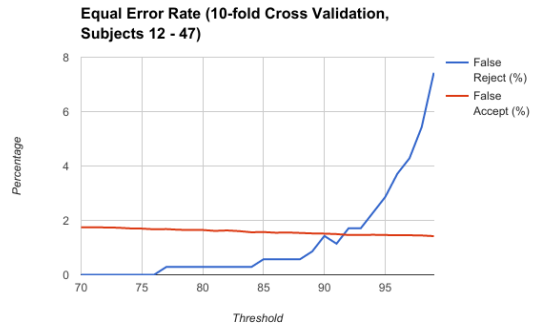
**Figure 14:** *RandomForest Parameter Tuning*




**(a)** *First 35 Subjects*




**(b)** *Middle 35 Subjects*




**(c)** *Last 35 Subjects*

**Figure 15:** *Majority Voting Results with the three overlapping sets of 35 subjects*

## vii. Training an Ensemble Classifier

Concrete success with the random forest algorithm led us to think about applying the ensemble learning concept to the system as well [5]. For any individual sample $S$, the classifier was constructed as follows.

1. Let $P$ be the set of feature points derived from the training partition of $S$. Since the training segment is 8 minutes long, and we are using 4 second windows, $P$ will contain approximately 120 points.

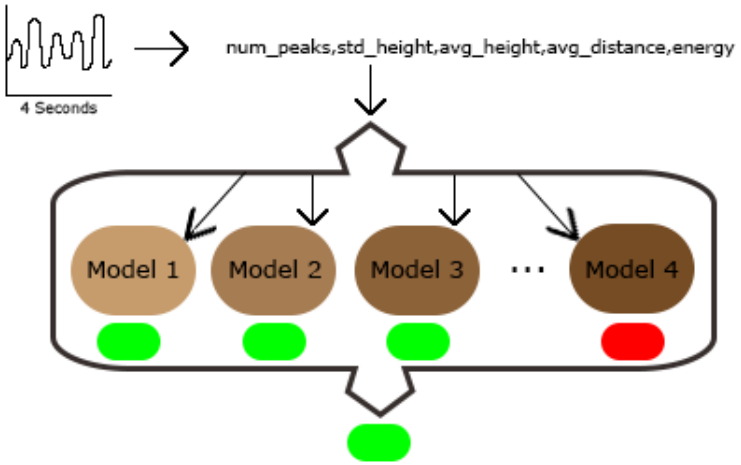



**Figure 16:** *Ensemble classifier example*

2. Let $N$ be the set of feature point derived from the training portion of all other samples not in the unseen set. The size of $N$ will be approximately 34 times the size of $P$, since there are 34 non-unseen samples that are not $S$.
3. Randomly shuffle $N$, and partition it into 34 equal size sets, $N_1, N_2, ..., N_{34}$.
4. Generate 34 Random forest classifiers $RF_1, RF_2, ..., RF_{34}$, where $RF_n$ is trained using $P$ as the positive set, and $N_n$ as the negative set.

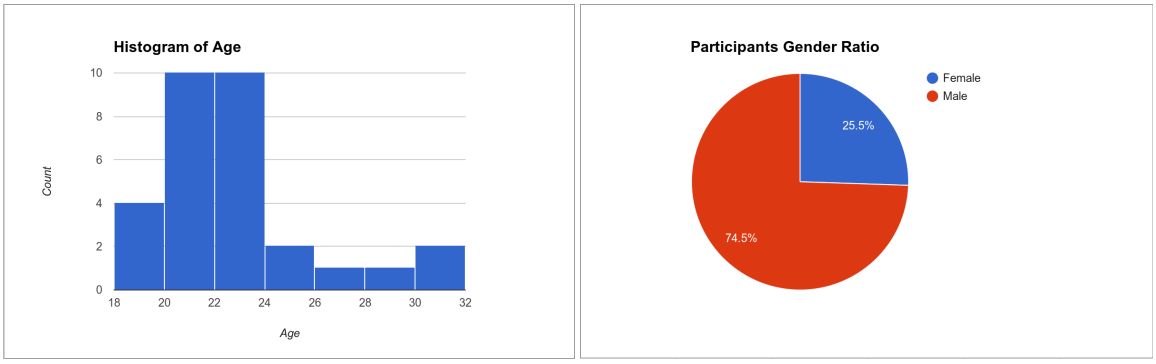
To determine if the ensemble classifier accepted or rejected a feature point, we queried each of the 34 models in turn. Each model responded with an acceptance or rejection. If a certain percentage or more of the models accepted the point, we considered it an acceptance, see **Figure 16**. We called this acceptance percentage the majority voting threshold. As the majority voting threshold increases, the FAR should go down, and the FRR should go up the FAR and FRR should meet at the point known as the equal error rate (EER). We decided to reserve 12 samples to use as totally unseen testing data, but we also wanted to make sure our accuracy was not affected by our choice of samples. Thus, we rotated which 35 samples we were using for training. We ran three trials. In the first trial, we trained models using the first 35 samples, then we used the middle 35 samples, lastly, we used the last 35 samples. The equal error rate of each trial is shown in **Figure 15**, located at 83%, 92%, and 92%, respectively. Averaging these numbers, we selected the optimum majority threshold at 89%

## VII. Experimentation

Ensuring our system is scalable and secure with higher amounts of subjects was extremely important, and as such our goal was to collect the data from as many people as possible, ideally with a wide range of ages and mix of genders. The final experimental setup involved collecting 10 minute samples from 47 people (12 female, 35 male, see **Figure 17b**). The average age was 21.8 with a standard deviation of 2.53 (see **Figure 17a**). The following steps describe our exact experimental procedure and setup, depicted in **Figure 7**.

1. Consent Form - the subject is greeted and given the IRB-approved consent form to read over, (See **Appendix B** for form).
2. Verbal Explanation - the experimenter gives a summary of the research and what the subject is helping with.
3. Questions & Consent - the subject is given a chance to fully read over the consent form and ask any questions. He/She is also expected to either give consent to the study or express concern and stop at this stage.
4. Relax - the subject is given time to adjust and get comfortable in their seating and procure a phone/laptop/book if they choose to do so.
5. Posture - the experimenter explains the reason for keeping a certain posture, sitting upright with elbows on desk, and informs them to keep as still as possible and not to talk or look around.
6. Start - once the subject is ready they are informed the study is starting and data collection, explained in System Model, is started along with a ten minute timer.
7. Check - Ensure subject is sitting in proper posture and not fidgeting.
8. Stop - once 10 minutes are up the collection is stopped and the subject is informed they are free to ask any questions or leave.
9. Store - the data is stored in a dedicated directory with an incrementing numerical ID.

We selected a Google Glass as the specific HMD for this experiment (see **Appendix A**), largely due to availability. The data communication process between the Google Glass and the Laptop is as explained in the System Model section. The main issue we encountered with the experimental setup is the length of the study combined with minimal activity leading to boredom. Each subject was allowed to read or watch video on a smartphone device to alleviate boredom, and by extension, avoid fidgeting. A sample setup can be seen in **Figure 7**. The goal of this setup was to mirror projected real world use as closely as possible, while still maintaining standardization from subject to subject and reproducibility in the future.




**(a)** *Age distribution of experiment participants* **(b)** *Gender breakdown between participants*

**Figure 17:** *Experiment demographics*

# VIII. Metrics

We used the following metrics to assess the effectiveness of our system:

### True Accept, False Accept, True Reject, False Reject

True accept (TA) is the number of positive samples accepted by our authentication system. False accept (FA) is the number of negative samples accepted by our authentication system. True Reject (TR) is the number of negative samples rejected by our authentication system. False Reject is the number of positive samples rejected by our authentication system. A perfect system will have 0 FA and 0 FR.

### True Accept Rate

The true accept rate (TAR) represents the proportion of positive samples that our system classified successfully. E.g. the number of true accepts, divided by the total number of positive samples ($\frac{TA}{TA+FR}$). The value of the TAR ranges from 0 to 1, where a perfect system will have a TAR of 1. This number is also known as true positive rate (TPR) or recall.

### True Reject Rate

The true reject rate (TRR) is similar to TAR. It represents the proportion of negative samples classified successfully. It is defined as the number of true rejects divided by the total number of negative samples ($\frac{TR}{TR+FA}$). The value of the TRR ranges from 0 to 1, with a perfect system having a TRR of 1. This number can also be referred to as true negative rate (TNR) or specificity.

### False Accept Rate

False accept rate (FAR) is the complement of the true reject rate, this number represents the proportion of negative samples that were accepted by our authentication system. It is defined as $\frac{FA}{FA+TR}$ or $1 - TRR$. This is also known as false positive rate.

### False Reject Rate

False reject rate represents the proportion of positive samples classified as negative. It is defined as $\frac{FR}{FR+TA}$ or $1 - TAR$. Another term for this number is false negative rate (FNR)

### Equal Error Rate

Changing parameters in a machine learning model will often result in contrary motion of the FRR and FAR. Making a system more secure will result in fewer false accepts, but increase the number of false rejects, and vice versa. The equal error rate is defined as the point where $FAR = FRR$.

### Precision

Precision represents what proportion of acceptances our system made were accurate. In a perfect system, only positive samples are accepted, so the precison will be 1. Precision is defined as $\frac{TA}{TA+FA}$.

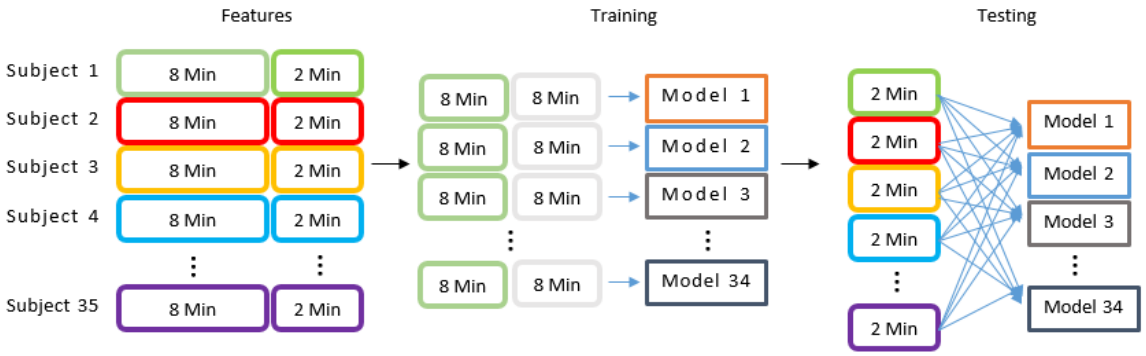



**Figure 18:** *Split testing (Subject 1 being the only who should be accepted)*

#### Balanced Accuracy

Balanced accuracy is defined as the ratio of correct classifications over all classifications. In other words, $\frac{TA+TR}{TA+TR+FA+FR}$. [5] It serves as a quick way to assess a system's performance, but is not a proper measure of true performance.

## IX. Results

### i. Final Parameters

The final system uses an ensemble classifier consisting of 34 random forests. Each forest contains 20 trees which each have a maximum of 2 features. When a new data point needs to be classified every random forest issues a classification. At least 89% of the classifiers in the ensemble have to accept the data point in order for it to be treated as an acceptance by our system.

### ii. Evaluation Methods

#### Split Testing

After collecting data, we partitioned it into the first 8 minutes and the last 2 minutes. Cross validation testing, described above, is conducted only using the first 8. Split testing works by using the 2 minute suffix as the testing set. We trained models using the first 8 minutes, and the tested the models with the 2 minute set to assess the true accuracy of the system.

#### Unseen Testing

Although we collected data from 47 subjects, so far, we have only been using 35 subjects worth of data. The last 12 subjects have been totally unseen to our system. Once we have built models for the first 35 subjects, we introduce the unseen data to the models. This lets us assess the ability of the system to interpret data from participants it has never seen before, mimicking the conditions under which a real system would operate. For unseen testing, all of the input data consists of negative samples, so there are no true acceptances or false rejections.

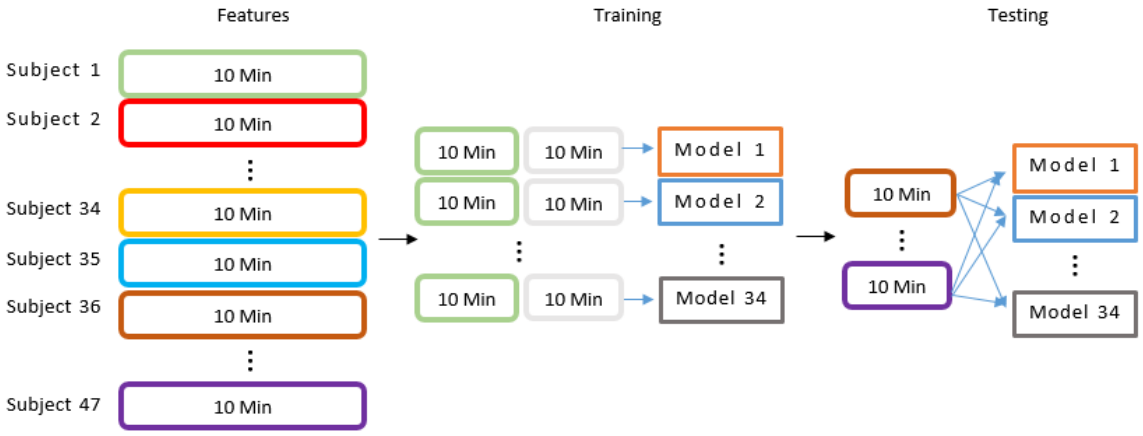



**Figure 19:** *Unseen testing (Subject 1 being the only one who should be accepted)*

## iii. Final Results

In order to ensure that all of the data belonging to the subjects had a chance to be in the seen section, we ran the 35 seen and 12 unseen test for three sections. The first 35 as seen, the middle 35 as seen and the last 35 as seen with the remaining 12 subjects in each being the unseen portion. In this way, we end with six separate tests: three split tests for each of the 35, as seen in **Table 3**, and three unseen tests for the 35, using the remaining 12 as the unseen, shown in **Table 4**. Collectively, these six tests shown that sampling the data in different ways has negligible effect on our results, and that the accuracy numbers hold.

| Metric | First 35 | Middle 35 | Last 35 |
|---|---|---|---|
| True Accept Rate (%) | 100 | 100 | 100 |
| False Reject Rate (%) | 0 | 0 | 0 |
| False Accept Rate (%) | 0.336 | 1.176 | 1.513 |
| True Reject Rate (%) | 99.663 | 98.824 | 98.487 |
| Balanced Accuracy (%) | 99.8328 | 99.412 | 99.244 |
| Recall (%) | 100 | 100 | 100 |
| Precision (%) | 99.665 | 98.838 | 98.510 |

**Table 3:** *Split testing results*

| Metric | First 35 | Middle 35 | Last 35 |
|---|---|---|---|
| Unseen False Accept Rate (%) | 2.619 | 4.762 | 3.571 |
| Unseen True Reject Rate (%) | 97.380 | 95.238 | 96.429 |

**Table 4:** *Unseen testing results*

### iv. Explanation of results

As can be seen from these results, overall, our models perform quite well. While there were a few false acceptances, these were relatively small given the size of the data set. As a result of this, though, it is difficult to see the current system deployed in a large-scale, high security context. It could perhaps find application as either a second factor means, or as a check of convenience. Due to its continuous natures, a system such as this can monitor a user's authenticated session after having authenticated via a more secure means; we believe that the security of this system should be more than sufficient for such an application.

## X. Discussion

While the results obtained from this experiment are certainly promising, there are a number of compromises made during the experimental procedure.

**Subject Diversity** All data was collected from volunteers on a college campus and as a result the diversity of our sample set is fairly low. A vast majority of participants were between 20 and 22. More work should be done to determine if our methods retain their effectiveness across different demographics in regard to age, gender along with health.

**Combining Sensor Components** When processing data, we selected a single sensor component by analyzing the frequency domain. Additionally, in testing this procedure, when selecting the component with the highest frequency response, we also tested creating a BCG waveform by aggregating all three axes of one sensor (L2 norm followed by z-score). However, we found that aggregating this data results in a waveform that is far noisier, and contains fewer easily findable defining features, thus, we discontinued this path. There may be a significant quantity of data lost when discarding 5 sensor components, and more work should go into finding a way to more effectively combine them.

**Different Positions/Postures** Our study primarily relies on subjects sitting since this is generally the most common position in society. The algorithm we designed should work regardless of the position assuming the data being collected is just as clean. However future studies should be done to find concrete evidence, one way or another, as to ensure a more solid backing for this authentication method. We recommend placing participants in a standing and lying down flat position to be more through.

**Changes in Heart Rate** Our data was collected from participants at rest. This means we did not test any subjects at anything other than resting heart rate. We are unsure if this system will work under conditions where the user has an elevated heart rate (e.x. After exercise).

**Changes over time** We have demonstrated that our authentication system achieves high accuracy rates for our data set. However, we have not found out if this accuracy holds up over time. In other words, once we build a model for someone, does that model continue to be reliable in the weeks and months afterwards? Depending on the results of this the system might need periodic re-enrollment as an individual ages potentially changing their heart beat patterns.

## XI. Future Work

The goal of this experiment was to establish the viability of using BCGs as an authentication medium able to uniquely identify individuals. However, this was conducted in a very limited context; subjects were restricted to a specific posture and prevented from moving. In reality, this is not a realistic expectation from users.

**User Movement Effects:** The resiliency of this system should be evaluated by allowing arbitrary movement by the user. Whether it be drastic head movement or simply talking to others near by, as without such resiliance it is impractical as a real world application.

**Shorter Training:** Training is a cumbersome practice that is a problem for any biometric authentication system, fingerprint or eyes or even voice. Expecting the user to train a model for 10 minutes however is impractical. Further work done with regards to how much this training window could be shrunk without significant impact on the accuracy of the model.

**Subject State Change:** Due to limited time and resources no work was done here to evaluate the impact of conditions that could temporarily influence the features of the user's cardiac profile, such as physical exercise. We predict that our features, being largely based upon the time domain, would rapidly deteriorate in accuracy if the same process was used. Additional work would need to be done with this process for the model to be able to recognize and adjust for this change. In short, the flexibility of this model with regard to its applicability in a variety of situations should be further explored.

**Different Body Position:** Assuming the algorithm and system hold up to changes in individuals, seeing if it is still as accurate and resilient on another part of the body would be interesting. A smart watch for example could employ such a technique to perform continuous authentication and act as an authentication hub for other portable devices.

## XII. Related Work

We examined multiple research papers which attempted to use Ballistocardiography as well as other Cardiac cycle related waveforms with Head Mounted Displays and other wearables. There are two papers that especially proved to be helpful in developing our method, BioGlass [7] and BioInsights [8]. These papers especially provided the initial procedures for generating a Ballistocardiogram waveform.

### i. BioGlass

BioGlass [7] conducted research and devised techniques for calculating the blood volume pulse (BVP) and respiratory waves from movement data observed with Google Glass. Data collection was performed with an experiment containing twelve participants, and movement data was collected from each. A medical device was used to generate a ground-truth BVP and respiratory wave which could be compared with the waves generated from the Glass to check the accuracy of the Glass. The movement data comes from the six axis accelerometer and gyroscope in the Glass, which records at 50Hz. The data is then interpolated to 256Hz to match the data collected with

the known to be accurate medical device, and used to generate the two waveforms. Summarily, the goal of this paper was to show that it is possible to generate physiological signals that are comparable to those gotten from intrusive medical devices using head mounted devices. Using these samples, the researchers were able to show definitive evidence proving that the signals generated from movement data showed a mean error of 0.83 BPM from the ground-truth generated from the medical device readings.

### ii. BioInsights

BioInsights [8] also followed some parts of the BioGlass paper but added on identification of the subjects. The researchers provided additional procedures for potential feature points in a BCG. This paper is also structured into three parts Data Collection, Cleaning and Analysis. We will once again summarize each of the sections and for more in-depth information the reader should read the original paper.

The collection procedure is similar to that performed in the BioGlass research. There are 12 participants and their movement data is collected either with a Google Glass head mounted display or Samsung Gear watch. Each participant has to record a minute of data while sitting, standing and lying down. Then they have to pedal a bike for one minute and then collect data again for each of the positions for a minute each. There was no ground truth device generating data known to be true. Both the Glass and Gear contain an Accelerometer and Gyroscope and the movement data from these sensors is collected for the participants. Once collected it is again important to clean the data and extract the waves we need while removing all other noises.

This paper involved simply extracting a BCG and none of the other data so the cleaning process is simplified and parts of the process originate from the BioGlass paper mentioned above. The first step is calculating the z-score for each point in the set of data for each axis. The next step is using an averaging filter with the window size set as 35 samples to detrend the data and remove unrelated body motions like respiration. Each of the axis is then passed through a Butterworth band-pass filter with cut-off frequencies of 4 and 11Hz to remove the unrelated frequencies from the data. Lastly using Fast Fourier Transform the axis with the most distinct peaks and troughs is selected as the BCG.

### iii. Existing HMD authentication methods

Currently, authenticating for a Google Glass requires the use of a desktop computer, laptop, or smartphone. The user enters the credentials for their google account into the MyGlass website or app. The app then generates a QR code, which can be scanned by the headset to log in. [1] Another popular HMD, the HTC vive, uses a different authentication mechanism, leveraging the Vive's movement-tracking paddles. First, the user must create an HTC account. When wearing the HMD, a virtual keyboard will appear, and the user will be prompted for their username and password. The user points the paddles at the virtual keyboard to type in their credentials. Both of these methods are cumbersome and time-consuming, forcing the user to spend a significant amount of effort to log in to their device. Additionally, once the devices are logged in, there is no mechanism for detecting if the authenticated user has been replaced with an adversary. We believe that HMD authentication should be fast, require minimal interaction from the user, and should be able to continuously verify the identity of the user.

## XIII. Conclusion

In this experiment, we have demonstrated with concrete results that any head mounted device with integrated accelerometer and gyroscope can act as a biometric authentication device without additional hardware. Given raw accelerometer and gyroscope readings from a head mounted display our system is able to construct a BCG representative of the wearer's heartbeat.

Given the BCG waveform, we are then able to compute features that we have shown to be highly unique between individuals. Utilizing ensemble learning implemented with multiple random forests, that are trained on distinct datasets, our system is able to classify the new waveform with high accuracy. Furthermore, by performing true holdout, using part of our dataset as unseen data along with performing cross validation we have shown the strength of this system as a general authentication system based on a inherent data source in all beings.

# Bibliography


[1] *Setting up Wi-Fi*.

[2] Glass diagram, May 2015.

[3] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[4] Ryan Dahl et al. Nodejs, 2009–.

[5] Mikel Galar, Alberto Fernandez, Edurne Barrenechea, Humberto Bustince, and Francisco Herrera. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(4):463–484, 2012.

[6] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[7] Javier Hernandez, Yin Li, James M Rehg, and Rosalind W Picard. Bioglass: Physiological parameter estimation using a head-mounted wearable device. In *Wireless Mobile Communication and Healthcare (Mobihealth), 2014 EAI 4th International Conference on*, pages 55–58. IEEE, 2014.

[8] Javier Hernandez, Daniel J McDuff, and Rosalind W Picard. Bioinsights: Extracting personal data from âĂIJstillâĂİ wearable motion sensors. In *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pages 1–6. IEEE, 2015.

[9] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[11] Eduardo Pinheiro, Octavian Postolache, and Pedro GirÃčo. Theory and developments in an unobtrusive cardiovascular system representation: Ballistocardiography. *The Open Biomedical Engineering Journal*, 4(1):201-216, May 2010.

[12] W. R. Scarborough, S. A. Talbot, J. R. Braunstein, M. B. Rappaport, W. Dock, W. R. Scarborough, W. F. Hamilton, J. E. Smith, J. L. Nickerson, S. A. Talbot, and et al. Proposals for ballistocardiographic nomenclature and conventions: Revised and extended: Report of committee on ballistocardiographic terminology. *Circulation*, 14(3):435-450, Sep 1956.

[13] Issac Starr, A J Rawson, H A Schroeder, and N R Joseph. Studies on the estimation of cardiac output in man, and of abnormalities in cardiac function, from the heart's recoil and the blood's impacts. *The American Journal of Physiology*, 127(1), Aug 1939.

[14] Scott Torberg. What's inside google glass. 2013.

# Appendix A

# Google Glass

Google Glass, released in 2013, is a lightweight, low-profile, head-mounted display. It runs Android 4.4 as its operating system on an OMAP 4430 system-on-a-chip with 2GB of memory. It includes an onboard Wi-Fi module capable of operating at 2.4 and 5.0 GHz. Finally, it contains both an accelerometer and gyroscope capable of sampling of at least 50Hz[14].

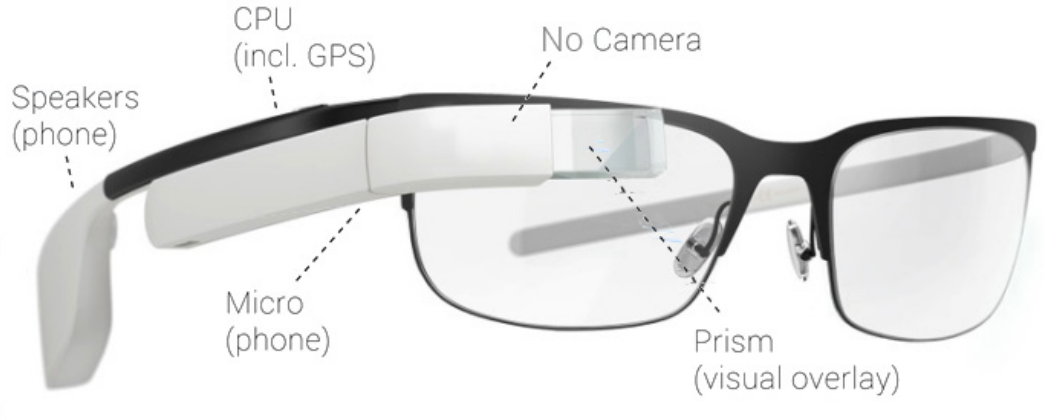



**Figure A.1:** *Google Glass HMD[2]*

# Appendix B

# IRB Form

# Informed Consent Agreement for Participation in a Research Study

**Investigator:** Krishna Kumar Venkatasubramanian

**Contact Information:** Contact Information: Fuller Lab 137, Computer Science Department, kven@wpi.edu. Tel: (508)831-6571

**Title of Research Study:** Authentication for wearable augmented-reality headsets

**Sponsor:** N/A

**Introduction:** You are being asked to participate in a Computer Security research study on developing a way to automatically identify the user (i.e., authenticate) of a wearable augmented-reality (AR) headset. Before you agree, however, you must be fully informed about the purpose of the study, the procedures to be followed, and any benefits, risks or discomfort that you may experience as a result of your participation. This form presents information about the study so that you may make a fully informed decision regarding your participation.

**Purpose of the study:** The purpose of the study is to determine whether *their cardiac rhythm characteristics extracted from minute head movements of user of an AR headset are unique enough to identify them*, automatically. If this hypothesis holds then the resultant technology will be like developing a ``password system'' for wearable AR technologies such as Google Glass, where the password entry process will be eliminated.

**Procedures to be followed:** You will be asked to sit and put on a Google Glass and make sure there is a comfortable fit. You will them be asked to look at a series of "cue-cards" on the Google Glass screen, which contain pictures. The goal is to count the number of dogs in the pictures. The purpose of the "cue-cards" is simply to keep the participants focused on a task for the duration of the data collection, so make it easier for us to collect noise-free readings. We plan to collect 5 minutes of data from each participant. The entire process should take about 5-6 minutes. During this process all you are expected to do is watch the "cue-cards" and sit without moving or fidgeting. We plan to record the following information during this data collection process:

- Your head movement through accelerometer and gyroscope sensors in the Google Glass device (from which your cardiac rhythm will be extracted).
- Your gender
- Your age
- Whether you are wearing contacts lenses.

If you wear glasses please remove them before putting on the Google Glass. If you wear contact lenses you can leave them in.

**Risks to study participants:** We do not anticipate any risks or adverse events wearing the Google Glass for such a short period of time. If however you are feeling uncomfortable at any stage of the data collection, please STOP IMMEDIATELY. At that time if you do not wish to continue, all your data will be erased immediately.

**Benefits to research participants and others:** The wearable AR headset technology will become more and more prevalent in the near future. They will be useful in a variety of applications from surveillance to services in public places like airports etc. It is therefore

important to ensure that the person wearing such technology is trustworthy. Being able to uniquely identify the wearer of the device is the first step in ensuring this property.

**Record keeping and confidentiality:** We do not plan to collect any personally identifiable information as part of this study. We will assign each participant of the study with a random id number. Any publication or presentation of the data will not identify you. All demographic data we collect will be reported in aggregate and will not single out an individual participant's response.

**Compensation or treatment in the event of injury:** You do not give up any of your legal rights by signing this statement. There is no potential medical risk or injury to the participants of this study. If the participant is injured there will be no compensation from anyone involved with the study and all medical expenses will be born by the participant or their insurer.

**For more information about this research or about the rights of research participants, or in case of research-related injury, contact:** Professor Krishna Venkatasubramanian, Tel. 508-831-6571, Email: kven@wpi.edu. WPI IRB Chair, Professor Kent Rissmiller, Tel. 508-831-5019, Email: kjr@wpi.edu. WPI Compliance Officer, Jon Bartleson, Tel. 508-831-5725, Email: jonb@wpi.edu.

**Your participation in this research is voluntary.** Your refusal to participate will not result in any penalty to you or any loss of benefits to which you may otherwise be entitled. You may decide to stop participating in the research at any time without penalty or loss of other benefits. The project investigators retain the right to cancel or postpone the experimental procedures at any time they see fit.

**By signing below,** you acknowledge that you have been informed about and consent to be a participant in the study described above. Make sure that your questions are answered to your satisfaction before signing. You are entitled to retain a copy of this consent agreement.

_____ Date: _____
Study Participant Signature

_____
Study Participant Name (Please print)

_____ Date: _____
Signature of Person who explained this study

**Special Exceptions:** Under certain circumstances**,** an IRB may approve a consent procedure, which differs from some of the elements of informed consent set forth above. Before doing so, however, the IRB must make findings regarding the research justification for different procedures (i.e. a waiver of some of the informed consent requirements must be necessary for the research is

to be "practicably carried out.") The IRB must also find that the research involves "no more than minimal risk to the subjects." Other requirements are found at 45 C.F.R. §46.116.