Using Heart Rate as a Method of Identification

Gabe Montague
Harvard College
(+1) 617-308-6726
gmontague@college.harvard.edu

ABSTRACT

In this paper, we discuss motivations and concerns surrounding identification of individuals based on the heart rate biometric. We also offer a simple distance-based algorithm that achieves, at least at a small scale, reasonable accuracy. The algorithm is able to identify individuals with an accuracy of over 50% given a recent knowledge base. We find that heart rate data falls into a daily and weekly periodicity, and that the most useful data for classification is the data closest in time to the sample that must be guessed.

General Terms

Algorithms, Measurement, Security, Legal Aspects.

Keywords

Biometrics, Health, Heart rate, Security, Identification, Privacy

1. INTRODUCTION

Wearable health-related technology has been on the rise in the past decade, as an increasing population of consumers are seeking commercial datification and analysis of their well-being. In 2016 alone, the number of wearable devices created experienced a growth of nearly 30 percent from its estimated total of 79 million in 2015.[1] By 2020, this number is predicted to skyrocket to 411 million devices, with wearable tech estimated to become a 34 billion dollar industry.[2] At the core of the existing and projected fitness wearable industries are devices and connected platforms that monitor health and fitness.

A large number of the emerging health and fitness systems, such as FITBIT and Apple Watch, incorporate the measurement and analysis of heartbeat data. Heartbeat data can be used most obviously to determine intensity and frequency of physical activity, but also hosts a wide variety of other analytical uses. Implicitly stored in a heartbeat signal is information that can be mined to determine sleep patterns, activity types, and individual levels of fitness, hydration, intoxication, fatigue, illness, stress, and mental health, among types of analyses.[3] In this way, heartbeat encapsulates both the short term day-to-day activities of an individual, as well as an individual's long term physiological state, and should thus be treated as information that is highly private to an individual.

Our goal in this research is to access the risk posed to an individual by their heartbeat signal falling into the hands of adversaries. In the following section we will formally identify the types of potential adversaries that could act to exploit this data, as well as possible means of exploitation. Our focus with the rest of the paper is to investigate possibility and risk of de-anonymization of heartbeat signals. Such de-anonymization would allow an adversary to listen or extract an anonymized heartbeat, then link the personal data extracted to an individual.

To quantitatively access this risk, we take the heartbeat signals of a team of ten athletes over the course of 92 days, and use it to build a simple classifier algorithm that successfully de-anonymization with a success rate of over 50 percent in short term cases. We hope to show that such de-anonymization allows for adversaries to link personal data in this form to the source individual, and should be viewed as an increasingly credible concern for the future privacy of individuals.

2. THREAT MODEL

We define an adversary in this context as any actor that could use an individual's heartbeat data to achieve ends that are contrary to the individual's interest.

2.1. Data Sources

An adversary may procure private heart rate data in any number of ways:

Firstly, an adversary could negotiate with the individual to exchange the data in return for benefits or rewards. For instance, companies may require their employees to wear fitness wearables and share the harvested health data, offering discounts on company health-insurance plans if the data indicates that the employee is healthy.[4] An adversary could also obtain heartbeat data that is made accessible online with inadequate security protections. In both these cases, the heartbeat signal and all of its derived data is directly tied to the individual. However, for the scope of this paper we will concern ourselves with eavesdropping cases in which the identity of the individual being mined is unknown at the outset, and the individual often does not know that they are being monitored.

In these cases, an adversary could obtain heartbeat data by monitoring streaming channels from the wearable itself. Although many modern wearables use Bluetooth wireless communication, a protocol that allows for encryption of streamed data, many device-makers are anecdotally known to make poor use of the protection capabilities made available to them, for example by using a constant trivial encryption key for all devices and communications. Additionally, the Bluetooth protocol itself comes with its own vulnerabilities; for instance a number of security holes were found with Bluetooth's recently updated protocol for Internet of Things devices: Bluetooth 4 low-energy.[5] Together, such vulnerabilities allow for easy eavesdropping on devices for listeners within a close-enough range of the victim.

Another interesting case is the possibility of an adversary collecting heartbeat data on an individual without any wearable at all. New technology developed at MIT's Computer Science and Artificial Intelligence Laboratory allow for heartbeat data to be extracted from stationary individuals from a video feed with a startling degree of accuracy. With this method, heartbeats are extracted from seated, standing, or sleeping individuals merely from small motions of the head resulting from blood circulation. The method is accurate enough to determine the number of milliseconds between heartbeats (determining HRV), and robust enough to measure heartbeats of individuals that are not facing the camera or wearing a mask.[6]

Using either of these previous two methodologies, an adversary could potentially obtain personal yet anonymized data in the form of a heartbeat signal. The question then becomes what incentives an adversary has for collecting this personal information, assuming he or she is able to link the data back to the individual.

2.2. Incentives

Due to the highly personal nature of information contained in a heartbeat signal, there are a number of reasons an adversary would be compelled to harvest the heartbeat data of individuals.

One of the largest incentives is for employers, for whom hiring employees in poor health could result in harm to the organization. The most productive employees are generally thought to be those that are mentally and physically healthy, and it is for this reason that many organizations sponsor organized athletic activity for employees. In cases where an employer provides healthcare coverage to its workers, corporate interest in the health status of employees is particularly strong.

Approximately 60 percent of United States residents rely on a healthcare plan provided by their employer, and this number is expected to continue to grow in coming years.[7] Although organizations are prohibited from discriminating against employees in poor health by federal law, companies may use personal health data in determining the amount to cover in individual healthcare plans.[7]

Use of healthcare data in determining healthcare coverage is a practice that is growing with the number of fitness wearables made available. As a result, modern companies in the United States are increasingly outsourcing health-status determination to third party companies, who market the scraping of employee health data to employers, especially as no federal law prohibits surveilling of employees in this way.[7]

Due to the highly personal nature of the knowledge of day-to-day activities such as sleeping, drinking, and physical activities, as well as the personal nature of health metrics provided heartbeat data, there exist any number of other uses of this derived information that are specific to the monitored individuals in question.

2.3. De-Anonymization

Given that in many ways, the most readily available streams of heartbeat data come from sources that do not reveal the identity of the eavesdropping victim, an adversary must put a name to the heartbeat data to make use of it. A number of factors, such as physical location and other circumstances of the collection, allow an adversary to narrow down possibilities for the identity of the victim, but as we find, there are a number of attributes in the data itself that indicate the individual.

The heartbeat data can be cross-validated in many ways, but our focus in this paper revolves around validation of identity based on previous known samples of heartbeat signals. If an adversary has obtained a sample of heart rate signals generated by known individuals, we find it then becomes easier to assign an identity to future heartbeat signals. This would enable adversaries to collect heart rate data by eavesdropping in bulk, then labelling each signal with an identity based on an existing knowledge base to make use of the harvested data.

Using high-resolution electric-based heartbeat data in the form of an electrocardiogram (ECG), such de-anonymization has already been shown to be possible with a high degree of accuracy among the entire human population. Given knowledge of any individual's ECG, existing algorithms are able to identify the individual at a later time from another ECG.[8] However, ECG's require more complex devices to measure than raw heartbeat-pulse data, and are therefore do not pose a significant privacy concern.[8] As we have seen, the heartbeat data more readily available for exploitation to an adversary is more likely to be a courser heartbeat dataset emanated from devices, which does not include information about the body's electrical signals, and likely is not sending beat-by-beat information, but information in the form of minute-by-minute heart rate data.

3. DATASET

To investigate the possibility of de-anonymization from course heartbeat data available to eavesdroppers, a dataset was obtained for experimentation of the minute-by-minute heart rates of ten athletes on an athletic team. Each athlete used a wearable with an optical wrist sensor of moderate accuracy to collect data, and each athletes was instructed to wear 24 hours a day for a period of 92 days.

3.1. Limitations

It is important to note that this dataset is comprised of individuals that participated in the same activities during over three hours each day of the week, and that the athletes observed practiced many of the same day-to-day activities, which is hypothesized to have reduced the accuracy of differentiation.

Also important is the observation that, due to limitations with charging the devices, not all of the athletes were successful in producing a steady data stream for the entire period of 92 days. In fact, all but one of the athletes were found to have significant holes in their data – stretches of many hours in which no heartbeat was logged to the system. The daily percentages of desired use for each athlete are plotted in Figure 1. However, these limitations mirror real scenarios of consumer data-collection in which an adversary is targeting a group.

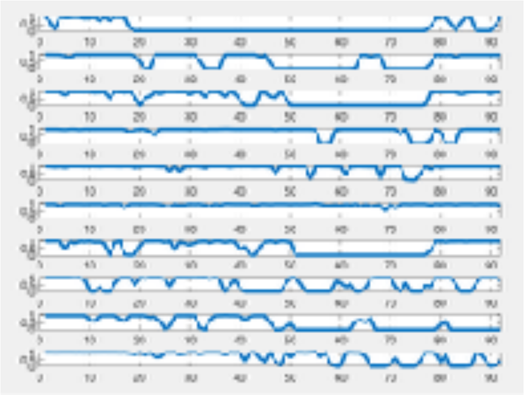

Figure 1. Plot of daily compliance (0 to 1) over each day.

3.2. Preprocessing

In preprocessing data for use in our identification algorithm, we proceeded with the ideology of making features, such as exercise and sleeping, as easily identifiable as possible. The following problems were addressed in preprocessing:

- Noise: Signals from optical sensors carry noise, and prevent features such as periods of elevated heart rate from being fully consistent.
- Holes: Holes in the data complete erase features, or disrupt their form.

To address the problem of noise, multilevel wavelet denoising was used. This methodology yielded strong results in a similar experiment conducted by Miro Enev et. al. in their publication *Automobile Driver Fingerprinting*, allowing differentiation between drivers.[9] When comparing our own algorithm's results (section 4) from raw data versus denoised data, it was found that the denoised data yielded stronger results as well, and was effective over standard median-filtering of the heart rate signal. It was found that decomposition using the Haar wavelet yielded the most consistent denoising of data.

To address the problem of holes, linear interpolation was used, and was found to have similarly positive results on the outcome of the algorithm.

TODO: Show preprocessing.

4. ALGORITHM

We sought in this study to construct a simple algorithm for heart rate data de-anonymization among small groups. We hoped our algorithm would:

- Demonstrate possibilities for de-anonymization of an arbitrary heart rate signal
- Be relatively simple and provide a building block for more robust machine-learning algorithms for identification
- Be extensible to larger datasets

4.1. Motivation

At the heart of many signal-based recognition algorithms is a matching function that compares two time-series data-streams. For instance, time series matching and comparison serves as a basis for the modern music-recognition software Shazam.[10] However, whereas music-recognition algorithms match noisy measured signals to a constant known signal, no part of an individual's daily heart rate signal is constant over all windows of time. To mitigate this problem with heart rate matching the following hypotheses were leveraged:

1. Most if not all individuals follow daily routines. Therefore, an individual's heart rate signal, although not identical from day-to-day, nonetheless is marginally periodic with a period of one day.
2. Many individuals follow weekly routines. Therefore, an individual's heart rate signal will be marginally periodic with a period of one week.

Given that these two hypotheses are correct, we find ourselves able to employ a similar algorithm to music-recognition scenarios:

- A noisy sample signal from an individual is observed, with a timestamp
- The signal is compared to a database of previously observed signals, comparing in particular to parts that correspond time-wise to the same part of weekly and daily cycles.
- The database signal that most closely matches the sample signal in this way is the best candidate for a match.

In particular, this cycle-alignment matching means that if a sample is collected on a Wednesday at 2 PM, we match this signal to other Wednesdays at 2 PM we have collected in our preexisting database.

4.2. Distance Metric

To test these two hypotheses, a distance metric is needed to compare windows of heart rate signals. Per recommendations of Joan Serrá et. al. in *An Empirical Evaluation of Similarity Measures for Time Series Classification,* we use the dynamic time warping (DTW) distance metric in comparing the signals in our algorithm.[11]

Essential to the effectiveness of DTW as a distance metric is proper normalization of compared features.[12] However, leaving data unnormalized has the advantage that individuals with higher or lower average heart rates are further differentiated by the DTW algorithm. Between normalization by day, normalization by athlete, and no normalization in comparisons where all considered for use in the distance metric. To test each option, a student's T-test was conducted on the two periodicity hypotheses mentioned before for each type of normalization. The T-test was conducted upon windows of the length of a single day as samples. We selected day-by-day normalization, as it yielded the best results in the T-test (the lowest p value). Comparisons of normalizations are shown in Table 1 and Table 2. We were unable to effect the weekly hypothesis results by varying the normalization type. In general, we found that although each test was far from statistical significance, our distance metric gave evidence for both hypotheses.

Table 1. Week periodicity validation based on type of normalization

Normalization	By day	By athlete	None
μ aligned	353.0456	353.0456	353.0456
μ non-aligned	380.6239	380.6239	380.6239
σ^2 aligned	1.3361E+04	1.3361E+04	1.3361E+04
σ^2 nonaligned	1.3000E+04	1.3000E+04	1.3000E+04
p value	0.4743	0.4743	0.4743

Table 2. Day periodicity validation based on type of normalization

Normalization	By day	By athlete	None
μ aligned	421.0390	382.7449	1.0079E+04
μ non-aligned	461.0029	417.4920	1.2657E+04
σ^2 aligned	1.1509E+04	1.2154E+04	8.1403E+06
σ^2 nonaligned	1.1347E+04	1.1792E+04	1.3621E+07
p value	0.4414 (lowest)	0.4516	0.4951

Unsurprisingly, we found greater evidence for daily periodicity than weekly periodicity in the individuals, with a difference of roughly 40 total BPM in the average distances between aligned and nonaligned windows.

Finally the DTW algorithm operates by "warping" signal features to move them in time to align with paired features. The maximal warping duration was chosen arbitrarily to be 6 hours. From observing comparisons made, it is unlikely that time-warping at such a long duration would occur naturally within the algorithm, so this decision is unlikely to have effected the results, and was chosen for considerations of speed.

4.3. Algorithm

Given this distance metric between windows, we designed the simple algorithm based on the points listed in section 4.1. The algorithm takes as input a 24-hour window of signal data and classifies it as belonging one of the ten potential athletes. It makes this decision with the help of a database comprised of a few weeks-worth of previous heart rate signals and their attached names. Given a sample it works as follows:

1. For each athlete:
    a. For each window of the athlete, compute the distance to the sample window and add it to the athlete's score. If the window is week-aligned with the sample (same day-of-week), decrease the weight of the contribution to the score from 1.0 to 0.8.
    b. Keep track of windows that are empty or too sparsely populated with heart rate data, and at the end of the athlete's score tallying, adjust the score so that each missing day contributed the mean per complete day to the score.
2. Select the athlete with the lowest score. The athlete with the second lowest score is the second guess, the athlete with the third lowest score the third guess, etc.

Input samples that were empty or sparse of data below a threshold were discarded in the classification.

4.4. Potential improvements

There are a number of shortcomings in the algorithm used deriving from its simplicity that can be spotted without any knowledge of its performance.

Firstly, the algorithm as developed contains a number of arbitrary choices that should be machine-learned in future replications. Particularly, the decision to deal with data-less days as contributing the mean, as well as the weekday-alignment weight of 0.8 are likely to be suboptimal.

Apart from that there are a number of other likely hypotheses that can be weighed into the algorithm that would likely improve its performance. DTW performs better when the ends of a signal are weighted down, which is not implemented in this version.[11] It is also likely that the contributions of windows should be weighted down based on how far time-wise they are from the sample, as daily and weekly patterns are likely to change over time, even at the heart rate level.

In either case, this algorithm can likely be replaced by a time-based neural network or a number of parallel support vector machines, that incorporate a larger number of factors. Miro Enev et. al. demonstrate the effectiveness of these more sophisticated approaches in their fingerprinting of driver signals.[9]

5. Results

The algorithm was tried under four circumstances, varying the size of the knowledge base, as well as varying the length of time between the samples and knowledge base. The knowledge base size was tested at 25 days and 35 days in length, and the gap between sample and knowledge base was altered between a single day and ten days. Under these conditions, random samples were selected from the week following the gap, and the classification carried out 500 times. The accuracy of the classification for each scenario is given by Table 3.

Table 1. Accuracy of classification in each situation

	Small base	Large base
Close gap	63.0%	57.7%
Far gap	52.8%	57.4%

Second, third, etc. guess were also stored for each trial. In the case that the first guess was incorrect, the algorithm usually guessed the individual upon the second guess. Plotting accuracy vs. number of guesses given we obtain the results given in Figure 2, Figure 3, Figure 4, and Figure 5.

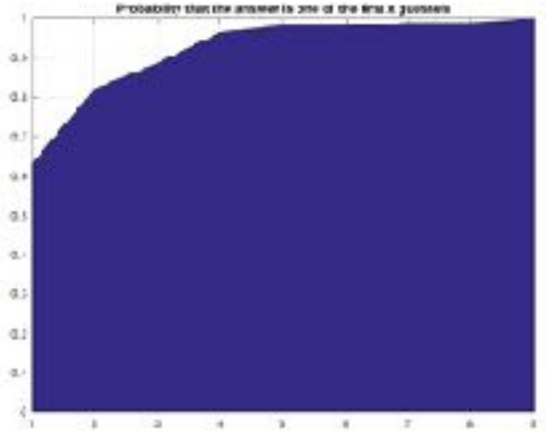

Figure 2. Accuracy for small base, close to sample

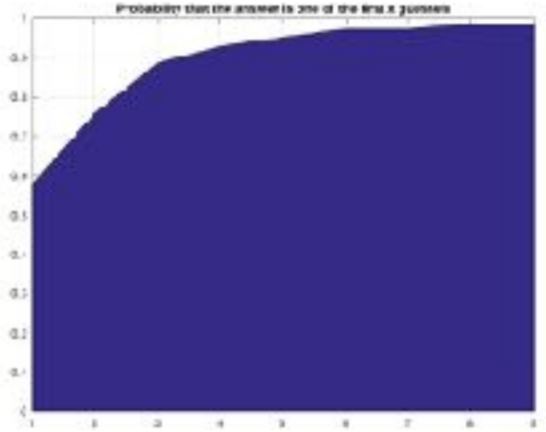

Figure 3. Accuracy for large base, close to sample

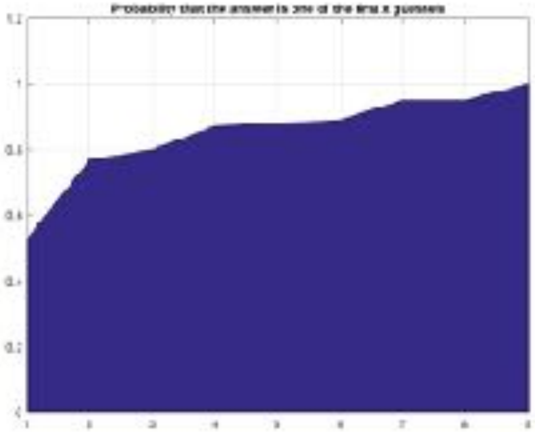

Figure 4. Accuracy for small base, far from sample

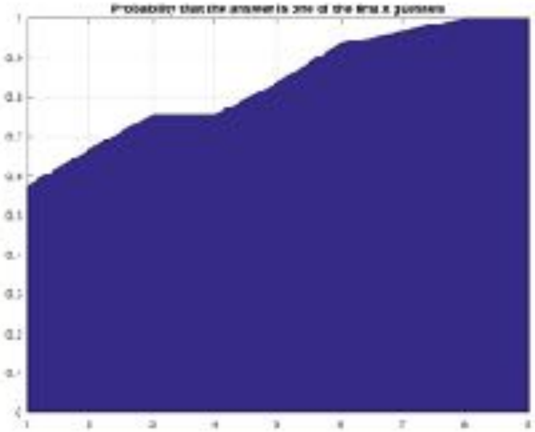

Figure 5. Accuracy for large base, far from sample

Figure 6 and Figure 7 show the comparison of accuracy when 24-hour sample windows with holes (missing data) were discarded from the classification results. In general this trend was adhered to by all four types of simulation.

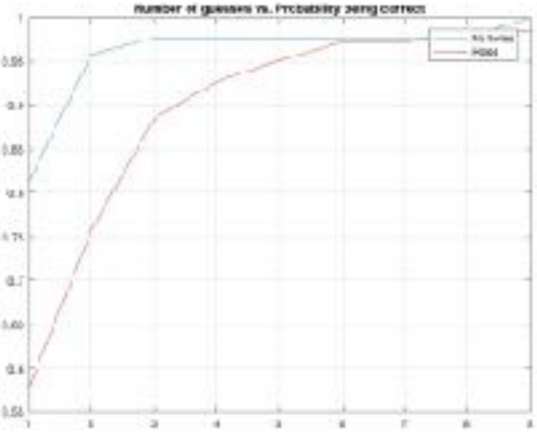

Figure 6. Holes for small base, close to sample

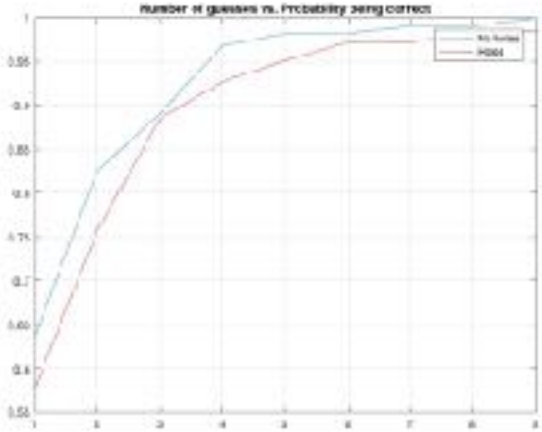

Figure 7. Holes for large base, far from sample

6. Discussion

We find from these results that in the case that our sample comes within seven days of our knowledge base, our simple algorithm is able to achieve a surprising degree of accuracy. However, as samples are taken further from the knowledge base, the accuracy of the algorithm decreases. This is likely because daily and weekly rhythms change over time, invalidating the knowledge base.

Perhaps the most puzzling result the fact that increasing the size of the knowledge base yields poorer results for a close gap, and stronger results for a far gap. However this is likely due to the fact that in the case of a small base and a close gap, the small base encapsulates all relevant cycles needed to distinguish upcoming samples. However if we increase the size of the base, older data is introduced, data encoding changed habits from months before that is not as relevant to deducing upcoming samples. In the case of a gap between knowledge base and samples, the knowledge base expanding leads to the inclusion of more relevant information that was otherwise lost in the gap.

This insight gives strength to our initial suspicion that the algorithm could be significantly strengthened by giving weight to the age of samples. It is clear from the data that older samples in a knowledge base are less valuable in classifying an incoming sample.

In general, we find that the accuracy of the algorithm is a weighted product of the following factors, for which the weights can be guessed from the graphs but not deduced with entire certainty:

1. The gap between the knowledge base and the sample
2. The number of samples in our knowledge base
3. The amount of data collected in our incoming sample that we must classify.

The first and the third are disproportional to the accuracy of the algorithm, while the second really depends on how the knowledge base is growing; if it receives newer, fresher data on individuals, then the accuracy of the algorithm will increase.

Finally the last and perhaps most important contributor to the accuracy of the data is the number of classification groups: the number of athletes being guessed from. Based on similarity to the work of Enev et. al., compression and optimization of this algorithm and similar ones is applicable at a large scale for identification. However, if the algorithm does not guess the correct individual from the data, and there is an increased number of individuals to choose from, it is likely that the true source of the data will become much harder to know judging from the guesses, as the algorithm run on 100 people will produce a ranking of 100, with likely only close to 60% chance that the first of the people is in fact the source of the sample.

However, also pointed out by Enev et. al., a number of data points surrounding the collection of the data can be leveraged to narrow down the list of candidate individuals to guess among. Given that the original context of the investigation was eavesdropping or recording by adversaries, an action that can only take place within a certain radius, real-life situations will only yield a fixed number of people within that radius that the heart rate signal could belong to. In this way we see that we have the beginning of a methodology that could potentially identify personal details of individuals on a massive scale.

7. Conclusion

With the growth of fitness wearables in the modern age, we find that an increased focus on privacy is in fact warranted. Vulnerabilities in systems allowing for eavesdropping of individual heart rates is now a reality. As we have seen, heartbeat data is an indicator of many more personal metrics than mere physical activity, some of which could be detrimental to an individual if the data fell into the hands of adversaries.

We see that with a simple comparison algorithm, it is indeed possible to sharply narrow down the possibilities for individuals that produced a heart rate signal, implying that in an increasing number of cases, unwitting users are emanating data that is inherently labelled with their identity, data that is highly personal and potentially detrimental to them in a number of scenarios.

We would like to explore how this algorithm could be further developed, and what steps we can take in the future towards ensuring a more private and secure system of health connectivity.

8. ACKNOWLEDGMENTS

Many thanks to Professor Almeida for his aid in motivating and helping along this research, and for providing a forum for discussion of this work. Thanks as well to our willing participants on the team that invested their time to measure their heart rate for such an extended period of time.

9. REFERENCES

1. (n.d.). IDC Forecasts Wearables Shipments to Reach 213.6 Million Units Worldwide in 2020 with Watches and Wristbands Driving Volume While Clothing and Eyewear Gain Traction. Retrieved December 16, 2016, from http://www.idc.com/getdoc.jsp?containerId=prUS41530816
2. Lamkin, P. (n.d.). Wearable Tech Market To Be Worth $34 Billion By 2020. Retrieved December 16, 2016, from http://www.forbes.com/sites/paullamkin/2016/02/17/wearable-tech-market-to-be-worth-34-billion-by-2020
3. Breslow, E. (2016). Heart Rate Variability – A Coach's Review of the Uses and Value of HRV Data in Athletes. *Whoop, Inc*. Retrieved December 16, 2016, from https://whoop.com/validation/hrv-overview.pdf.
4. Ajunwa, Ifeoma and Crawford, Kate and Schultz, Jason, Limitless Worker Surveillance (March 10, 2016). California Law Review, Vol. 105, No. 3, 2017, Forthcoming. Available at SSRN: https://ssrn.com/abstract=2746211
5. M. Ryan, "Bluetooth: With low energy comes low security," *Usenix*, 2013. [Online]. Available: https://www.usenix.org/conference/woot13/workshop-program/presentation/ryan
6. G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog*., 2013, pp. 3430–3437.
7. Ajunwa, Ifeoma and Crawford, Kate and Schultz, Jason, Limitless Worker Surveillance (March 10, 2016). California Law Review, Vol. 105, No. 3, 2017, Forthcoming. Available at SSRN: https://ssrn.com/abstract=2746211
8. Nymi Inc. (n.d.). HeartID Whitepaper. Retrieved December 16, 2016, from https://nymi.com/sites/default/files/HeartID-White-Paper.pdf
9. A. Wang. An Industrial Strength Audio Search Algorithm. In *Proc. ISMIR, Baltimore, USA*, 2003.
10. Abdullah Mueen, Eamonn J. Keogh: Extracting Optimal Performance from Dynamic Time Warping. KDD 2016: 2129-2130
11. M. Enev, A. Takakuwa, K. Koscher, and T. Kohno, "Automobile driver fingerprinting." *PoPETs*, vol. 2016, no. 1, pp. 34–50, 2016.

10. APPENDIX

The code used for this research can be found at https://github.com/montaguegabe/heart-rate-research. Sample data is not included in the repository.