Whose Move is it Anyway? : Authenticating Smart Wearable Devices Using Unique Human Body Movement Patterns

Sugang Li[†], Ashwin Ashok[*], Yanyong Zhang[†], Chenren Xu[*], Macro Gruteser[†]

[†]WINLAB, Rutgers University, North Brunswick,NJ, USA
[*]Carnegie Mellon University, Pittsburgh, PA, USA

ABSTRACT

The recent years have seen a significant growth in popularity of smart wearable devices. This growth can be attributed to the advances in hardware miniaturization technology as well as economically affordable and energy efficient sensing and computing. While size, energy and cost constraints remain key motives for improvements in wearable computers' design, the aspect of user authentication has received relatively less attention. Wearable devices often collect and store sensitive data about users, and thus there is an obvious need to authenticate the right user to the device. Solutions today primarily rely on indirect authentication mechanisms through the user's smartphone, which can be cumbersome and less secure. Biometric based solutions, though very effective, however, are subject to the availability of the specific sensors in the wearable unit. In this paper, we propose to authenticate wearable devices to users based on their unique behavioral patterns. In particular, we prototype an authentication system, dubbed *Headbanger* for wearable devices by monitoring user's unique head-movement patterns in response to an external audio stimulus. Using a personal imaging device as a running example and through extensive experimental evaluation over multiple users we show that our mechanism can authenticate users with an acceptance rate greater than 95% while keeping the average false rejection rate below 5%.

1. INTRODUCTION

We live in a world that has seen a generation of technological revolutions; from wired to wireless communications, immovable to mobile machines, large sized to hand-held devices. Today, we are witnessing what can be deemed as the next phase of mobile revolution through *wearable computers*. Research in wearable computers can be dated back to as early as 1980s when Steve Mann developed a prototype heads-up-display goggles []. Thanks to the advances in hardware miniaturization technology, cheap sensors/processor chips, and low-power sensing/computing, today, wearables are available off-the-shelf and have almost become an integral part of human lives [?, ?, ?].

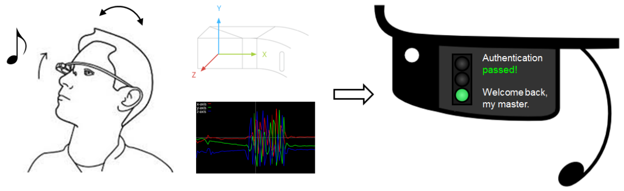


Figure 1: Illustration of Headbanger. The head-worn device authenticates the right user based on signatures generated from head-movement patterns in response to an audio snapshot played on the device.

The surge in wearables is clearly seen commercially through the array of smart wearable devices seeping fast into the market. Research on wearables has actively picked up and is progressing over the day. Keeping in mind the resource limitations on wearable computers research so far has primarily been addressing three key optimization parameters in design: size, energy and cost. With the onset of proliferation of wearable devices, in the recent times the aspect of security and privacy have also been adding key concerns to wearable computer usage. A solution for safeguarding the security and privacy of user's data on the wearable device is only effective as long as the device itself is authenticated to the right user/owner.

Authentication Challenge. Authentication on most commercially available wearable devices today [?, ?] relies on an indirect mechanism, where users can login to their wearables through their phones. This requires the wearable device to be registered and paired to the user's mobile device, which makes it inconvenient as the user has to carry both the devices. The security of this approach is also in question as it increases the chance of hacking into both the devices if either of the devices are lost or stolen. Some devices including Google Glass [?] and FitBit's health tracker [?], pair the device to the users email account instead of the phone for user's convenience; however, it does not add any security benefit. Though wearable devices today almost contain the same suite of sensors as a smartphone, the computing capacity and battery lifetime of wearables are far less comparable. This implies that, translating the same authentication solution from a phone to the wearable device, to enable direct authentication, is not only undesirable but also impractical. *Thus, the need for the day is a simple, low-power, and accurate direct authentication solution.*

There have also been a few specific point solutions that propose to directly authenticate the user to the device using biometric signatures [**?**, **?**]. However, collecting and recognizing these biometrics is subject to the availability of the sensing hardware and the computing capability on the wearable units, hence unrealistic in most cases. The recent years have also seen a significant interest in using behavioral characteristics of humans as biometric signatures for mobile phone authentication [30, 26, 24, 21, 5, 9]. For example, human walking gait, arm swings, typing patterns, body pulse beats, eye-blinks have all been found to be distinctive in human beings. The main advantage of using behavioral biometrics is that the signatures can be generated from raw data of basic sensor that are inbuilt in mobile phones such as motion sensors, camera, microphones etc. However, what behavioral biometric remains the best suited for a wearable device and if such a biometric can be captured by the wearable device lays a fundamental question for a system that may adopt this approach.

Head-movement based authentication. In this paper, we propose to authenticate wearable devices to users based on their unique behavioral patterns. Keeping in mind the feasibility of implementation and the availability of the sensor on the wearable device, we design a system that authenticates users to their device directly through their head-movement patterns. In particular, we prototype an authentication system, dubbed *Headbanger*, that generates a signature from user's head-movements that serves as the behavioral biometric for authentication. To ensure that the user is proactive in making head-movements we stimulate the process by playing a short duration audio track with fast beats. The user in response to the rhythm and beats makes head-movements that are captured by the accelerometer and processed to generate and authenticate the user's unique biometric signature. While we use a Google Glass a running example for the wearable device, our design can be applied to other head-worn gadgets and any system that can record head-movements through motion sensing. Our choice for using head movements is motivated by the fact that head-worn wearables are becoming very common today and such devices are already equipped with motion sensors; for example, personal imaging and heads-up display devices, gaming headsets, AI devices.

In summary, the key contributions of this paper are:

1. We propose a technique for direct authentication to wearable devices using head-movement patterns. We show that user's head-movement patterns contain unique signatures that when inferred correctly can be used as valid behavioral biometrics for authentication.
2. We design an authentication system, *Headbanger*, that records, processes, generates unique signatures, and classifies head-movement patterns of users based on the accelerometer (inbuilt on the wearable device) sensor readings.
3. We implement a data collection application on Google Glass that plays a fast beat music (preloaded) through the Glass's speakers and simultaneously records and filters accelerometer sensor data. We use this data collection app in our experiments to evaluate the system.
4. Through extensive experiments involving multiple users[1] and over different system design parameters we have shown that head-movement patterns can be used as a valid behavioral biometric. to effectively identify a smart wearable device user, with average false acceptance rate of ***% and average false rejection rate of ***%.
5. We have conducted detailed validation of using head-movement patterns as a biometric characteristic by collecting and analyzing data from multiple users. We will make these data publicly available, which may facilitate other user behavior studies.

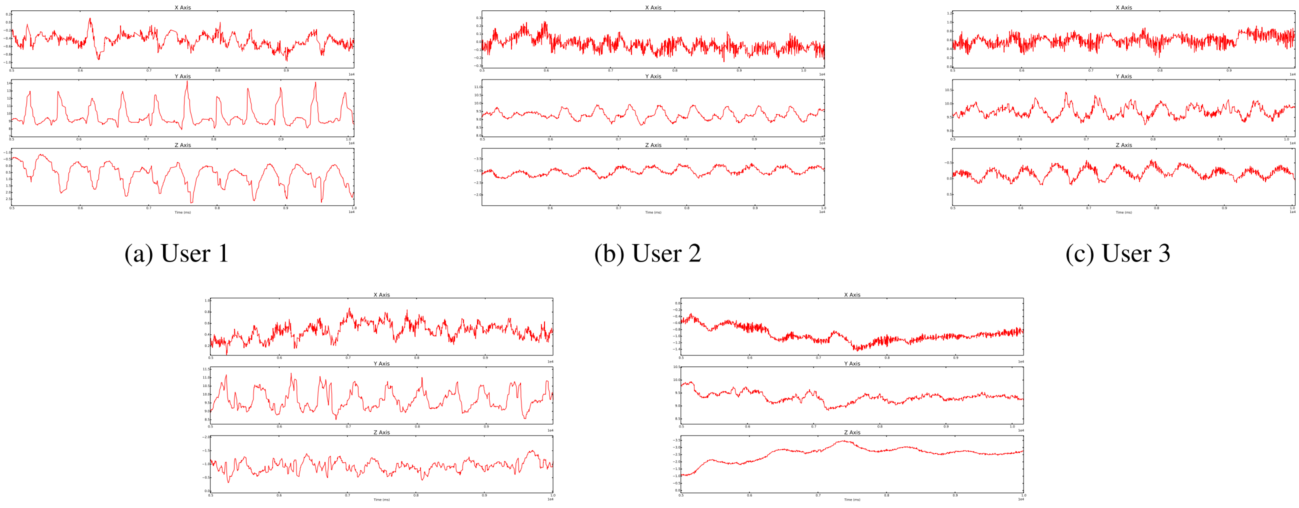
In the sections to follow, we will discuss the background of wearable device authentication in section 2 and details of our proposed system design in section 3. We evaluate the system in section 4 and follow up with discussions and conclusions in sections 5 and 7, respectively.

2. BACKGROUND ON WEARABLE DEVICE AUTHENTICATION

Authentication mechanisms for wearable devices can broadly be divided into two categories: (i) *Direct* authentication, where the users can directly authenticate themselves to their wearable device using the input/output interface and/or using signatures generated from the sensors available on the device, and (ii) *Indirect* authentication, where a secondary device – typically the user's smartphone – is used as a medium for authentication. Today's commercially available wearable devices predominantly use the latter approach where users login to their wearable devices through their smartphone – using a PIN or an email account. Unlike the indirect approaches, that require wearable device is registered to the smartphone and connected (wireless) to the wearable device, direct mechanisms can leverage from the inbuilt interfaces and sensors on the wearable device.

The fact that wearable devices relate significantly to "what we wear" on the human body, biometrics can play a key role for direct authentication to wearable devices. Biometrics allow a system to identify a user based upon "who you are"

[1] Our human user experiments were approved by Rutgers University IRB



(a) User 1

(b) User 2

(c) User 3

(d) User 4

(e) User 5

Figure 2: These plots show the raw accelerometer data in the time domain for five different users when they move their head in response to a music track wearing the same Google glass. The plots indicate that different users' head movement patterns appear distinctive from each other. The five users wore a Google Glass (in turns) and listened to a 10 second audio snapshot of a pop song.

(i.e., her physiology) instead of "what you have" (i.e., ID cards) or " what you remember" (i.e., passwords) [20, 25, 32]. Physiological biometrics such as DNA, ear shape, face, fingerprint, hand/finger geometry, iris, odor, palm-print, retinal scan, and voice, have been very effective and widely used in many prototype and commercial authentication systems. In addition, body shape such as body height, width, and body-part proportions can also be used as biometric cues to identify different people [8]. Even "soft" characteristics such as body weight and fat percentage have been considered as secondary biometrics for authentication purposes [2]. However, biometrics are not prominently used in wearable devices commercially available today. While there have been specific point commercial designs (Nymi [], BioNym []), it is unlikely that biometrics will become de-facto standard for authentication in wearable devices. This can be attributed to the fact that biometrics would require the specific hardware/sensor available on the wearable device. Also the overheads for physiological biometrics in wearable devices can be high, in both, cost for hardware as well as integration and computing.

An other approach to direct authentication is using behavioral biometrics where unique signatures from human behavior (subconscious or in response to external stimulus) provide cues for differentiating and authenticating users. For example, it has been shown that gait (e.g., stride length, the amount of arm swing) when the user is walking or running is a reliable identification cue, and irrespective of the environment [30]. Okumura et.al. [26] have shown that the human arm swing patterns can be used to create signatures to authenticate to their cell-phones. Monrose et.al. [24] show that keystroke rhythms, when users type on the keyboard, that include typing dynamics such as how long is a keystroke, how far is between consecutive strokes, and how is the pressure exerted on each key, can be used as a biometric to authenticate users. Similarly, mouse usage dynamics [21] and touchpad touching dynamics [5, 9] have also been shown to serve as potential biometrics.

In comparison to other means of authentication, behavioral biometric authentication can offer a more convenient (than physiological biometrics), and more secure (than indirect authentication) solution for wearable device authentication. With the increasing off-the-shelf availability and (almost) unlimited access to the sensors on the wearables, it has become possible to generate and/or infer unique behavioral signatures specific to users. We use these rationale as a motivation for our proposed design of a behavioral biometric based authentication that generates unique signatures from accelerometer signal patterns from user's head movements. We design an authentication system, dubbed *Headbanger*, for head-worn devices by monitoring user's unique head-movement patterns in response to an external audio stimulus.

2.1 Head-movement as a Biometric

According to [20], a human characteristic can be considered as biometric as long as it is *universal*, *distinctive*, *repeatable*, and *collectible*. With the advancements in head-worn wearable computer designs it is becoming easier for collecting head-movement patterns using the inbuilt accelerom-

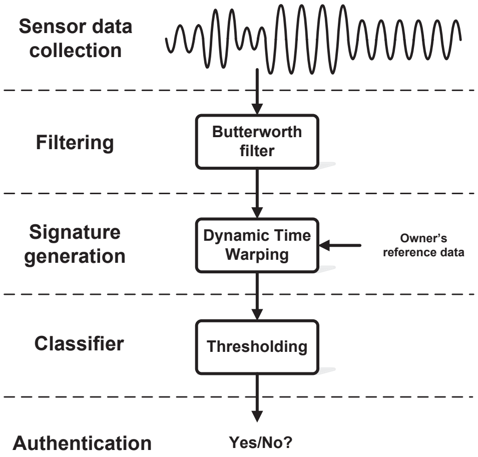


Figure 3: *Headbanger* system design flow

eters and motion sensors. Such sensors are available on almost all head-worn wearable devices available today, thus making head movements that are universally available *collectible* in all aspects. It has been shown [33] that most people move their body as a natural response to external rhythmic stimuli such as music – even at a very early age, infants respond to music and their movements speed up with the increasing rhythm speed. We observed that even adults naturally perform head-movements when listening to a fast beat audio track. Based on a preliminary analysis of the accelerometer signals from five Google Glass users, we also observed (see Figure 4 (a)-(e)) that these users *repeatedly* showed unique and *distinctive* head-movement patterns, when listening to the same music beats on the head-worn device. Motivated by this observation we hypothesize that head movements can be a good behavioral biometric characteristic to authenticate users to their smart glass. We next formally present the design of our system that utilizes head-movement patterns as behavioral biometric signature to authenticate smart glass users.

3. HEADBANGER SYSTEM DESIGN

In this work, we design a system, *Headbanger*, that will enable direct authentication of users to their smart-glass device using head-movements. We envision that our proposed system will be used as an authentication interface on the smart-glass wearable device. The system will run as a service in the device upon power-up, similar to the screen-lock in smartphones or the head-nod interface on Google Glass [**?**]. Upon usage, a short duration audio track will be played on the device, and the user makes head-movements in response to the audio. Our design is developed based on the idea that humans respond to music naturally through head movements, and that such movements are more significant and unique when the track contains fast beats and/or rhythm. *Headbanger* generates unique features from the head movements of a user, and uses them as a biometric signature for authenticating the right user to the device. The system will grant access only when the head-movement signature generated during the login attempt matches with the original user's signature.

As illustrated in Figure 3, the design of *Headbanger*involves the following key steps:

- *Sensor data collection*: *Headbanger* records the head-movements in the form of raw accelerometer signals using the inbuilt accelerometer sensor on the smart-glass device.
- *Filtering*: The accelerometer signals are filtered by applying a simple low-pass filter to remove records of extraneous motion.
- *Signature generation*: The accelerometer signals are processed through the dynamic-time warping (DTW) tool [**?**] to obtain a DTW feature that is treated as the unique signature for the user. A user can generate different signatures for different audio tracks. A training phase collects the set of signature for each user-audio pair.
- *Classification*: The signatures are classified to different users based on thresholding and machine learning strategies, and labeled to the corresponding user.
- *Authentication*: The head-movement signature generated during system operation is compared with the original user-audio pair signature, and the user is authenticated access if there is a plausible match.

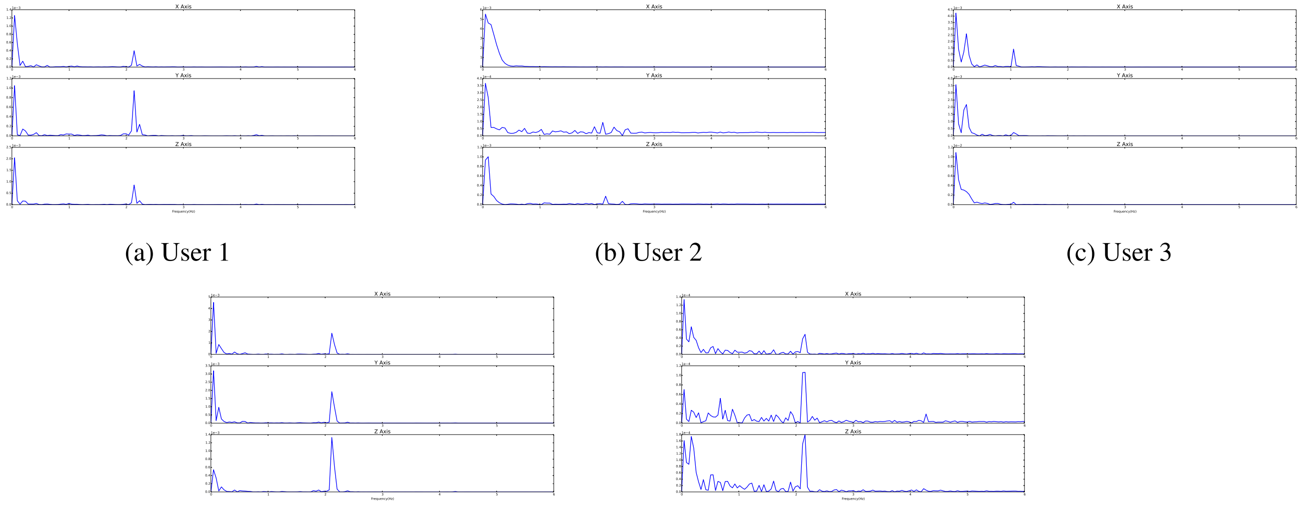
We will now discuss these design aspects in more detail.

3.1 Sensor Data Collection

The sensor data collection step involves the user wearing the head-worn device and making head-movements in response to the music beats played on the device for T seconds. In this duration, the raw accelerometer signals, from the inbuilt sensor, are collected at a sampling rate of r samples/sec; the default sampling rate on Google Glass is 200Hz. Each accelerometer signal sample, referred to as ACC from here-on, is a collection of accelerometer readings on the x,y, and z axis. The data collection unit stores each sample in a matrix with dimensionality $3 \times rT$ where each element corresponds to one signal point of ACC. We retain the duration T to be in the order of few seconds, as frequency of human head movements are, intuitively, typically in the order of few times per second magnitude.

3.2 Filtering

The ACC samples are cleaned-up from noise (accelerometer readings due to spurious movements such as vibrations) through a filtering stage. The filtering ensures that the head-movement signature generated from the accelerometer readings encompasses only head movements, and not any spurious signals caused due to vibration or shaking. From the



(a) User 1 (b) User 2 (c) User 3

(d) User 4 (e) User 5

Figure 4: Accelerometer data from 5 people, in the frequency domain.

frequency spectrum of each ACC sample, as shown in Figure **??** for five users, we can observe that the spectrum is significantly concentrated within 5Hz. We note that music tracks with high tempo, or fast beats, typically contain beats in the order of the order of hundred beats per minute. In particular, the high tempo music used in our experimentation was defined to have 94 beats per minute [23]. We infer that the head-movement, in response to the beats, will be of the same order. Hence, we hypothesize that the signal spectrum in [0,5] Hz range encompass human head movements; where 0 Hz can indicate that the head is steady still, and 5 Hz can correspond to a vigorous head-shake. We filter ACC samples using a low-pass digital Butterworth filter [**?**]. We set a relaxed cut-off frequency of 10Hz. Even with the relaxed cut-off, the filtering results in clean accelerometer samples with head-movement patterns that are prominent and detectable. Figures 5 (a)-(e) show the filtered results of the raw ACC samples shown in Figure 4; clearly, the filtered results are much more suitable for subsequent processing.

3.3 Signature generation

We generate a signature from the accelerometer signals using the dynamic-time warping (DTW) tool. DTW is generally used as a similarity matching tool for time-domain analysis of temporally varying signals. DTW compares a temporal signal with a reference signal over a certain time-window and yields a distance measure as the score. A low score (distance) implies that the test signal is in close match with the reference. We use the DTW principle to generate a signature for the head-movements from the accelerometer signal. We observed from our initial experiments that, users often start head movement at a different angle, but exhibit a consistent, and often periodic, patterns over time. Treating the ACC sample set from the first trial as a reference, and we apply the DTW algorithm on the successive ACC sample set to obtain a distance score vector, $\hat{d} = (d_x, d_y, d_z)$; the three elements in $\hat{d}$ denote the DTW distance score in the x, y, z axes, respectively. By computing the mean of the distance scores obtained for each ACC pair we generate the mean-value DTW distance, which is treated as the head-movement signature for that particular user and audio combination. For evaluation purposes, we conduct an elaborate training phase where N trials of head-movement exercise are done, to obtain $N-1$ distance vectors. In real usage of the application, the user would conduct only two trials of T duration each where trial 1 is treated as reference. In this way, the training can be done in-situ. The training data-set is updated upon each usage of the authentication interface.

3.4 Classification

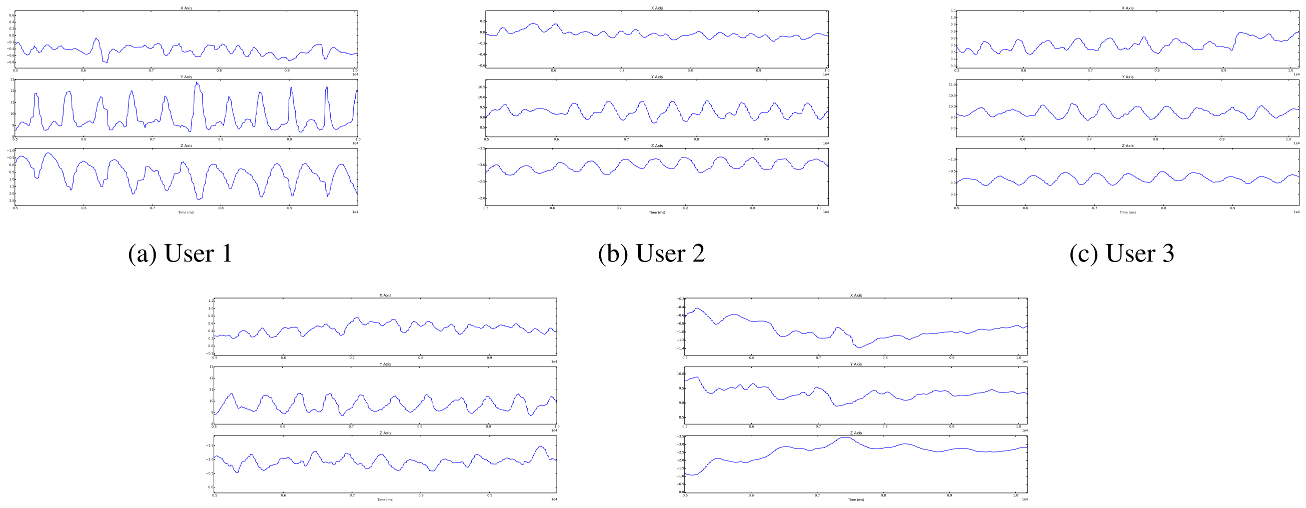
The classification step labels a test sample based upon the pre-established training samples, or trained signatures. In this work, we consider two classification mechanisms:

- *Supervised learning (SVM)*, which generates the trained signatures for each user based on a training set that includes the user's original signature as well as a number of other false signatures, that do not correspond to the user.

- *Adaptive thresholding*, which generates the trained signatures for each user based on the signatures generated from a select set of trials of the same user.

We will now discuss these two methods in more detail.

3.4.1 SVM-Based Classification



(a) User 1

(b) User 2

(c) User 3

(d) User 4

(e) User 5

Figure 5: Filtered accelerometer signals. Applied Butterworth filter of order 2 and cut-off frequency 10Hz.

The SVM based classification goes through the following steps:

1. Construct the training set by including both “true” samples, corresponding to the device owner, and “false” samples, corresponding to the signatures from random users.

2. Calculate the DTW vectors between any two true samples (with the resulting DTW vectors referred to as $DTW_{T\to T}$), as well as DTW vectors between any true sample and false sample (with the resulting DTW vectors referred to as $DTW_{T\to F}$).

3. Given a test sample S, calculate the DTW vectors between S and any true sample (with the resulting vectors referred to as $DTW_{S\to T}$) as well as DTW vectors between S and false samples (with the resulting vectors referred to as $DTW_{S\to F}$).

4. Input both, training DTW vectors (i.e., $DTW_{T\to T}$ & $DTW_{T\to F}$) and test DTW vectors (i.e., $DTW_{S\to T}$ & $DTW_{S\to F}$) to a binary SVM classifier that labels each signature with a 1 or 0 along with an identity corresponding to the user-audio pair; a 1 meaning the signature belongs to the right owner and 0 that it does not.

Optimized SVM. In order to minimize the computing overhead in the above process and make it suitable for running on a smart-glass, we conduct the SVM classification in an optimized fashion: instead of looking at every single training sample from a user, we choose a small set of representative samples and use them for classification. Here, we define a user’s representative samples as those that are the most similar to samples from other trials (of the same user). Let us suppose a user’s entire training set has N samples (from N trials). For every sample, we compute its distance to each of the $N-1$ samples using DTW. Then we choose the K samples that have the lowest K average distance to the rest of the training set. By setting an appropriate threshold K, we find the representative samples from the training set. We refer to these K samples as top K samples of the user. If we choose a small K value, say 3 as opposed to 20, the computing overhead can be significantly reduced. We conduct this optimization strategy keeping in mind the increasing latency of SVMs with training set size. We note to the reader that such an optimization is only effective when the unoptimized training data-set size is sufficiently large. Eqn. 1 shows how we calculate the top 1 sample from N samples from a user.

$$\begin{aligned} DTW_{\mu} &= \frac{\sum_{i\neq n} DTW_i(n)}{N} \\ DTW_{min} &= min\{DTW_{\mu 1}, ..., DTW_{\mu N}\} \end{aligned} \quad (1)$$

3.4.2 Adaptive thresholding-Based Classification

Through empirical trials we observed that the SVM classification may not be very robust against false training signatures that come very close to the true samples. While finding optimal binary SVM classifier parameters or using a multi-class SVM [**?**] are probable solutions, we felt that these approaches will add significant overheads to the already complex SVM. As an alternative, we developed a simple yet effective classification scheme based on adaptive thresholds. We highlight the aspects of the adaptive thresholding procedure as follows:

1. Given N true training samples, identify the top K samples (representative set), and calculate their DTW vectors to the rest of the samples in the training set to ob-

tain $(N-1)K$ distance vectors. Call this resultant vector as top-K distance vector.

2. Compute the 3 element (x,y,z axis) mean μ vector and standard deviation σ vector of the top-K distance vector. Define the "true" range as $[\mu - m\sigma, \mu + m\sigma]$, where m is a design parameter that will be fixed by the designer. The samples outside this range are labeled as "false". We refer to this range $[\mu - m\sigma, \mu + m\sigma]$ as the threshold in our classifier design.

3.5 Authentication

The authentication step results in a binary output that corresponds to either allowing or disallowing the user to unlock the device. In this paper, we make two reasonable assumptions pertaining to authentication on the smart-glass device: (i) homogeneity of accelerometer sensors for all smart-glasses (from a particular vendor), and (ii) that the device is registered to the user with an associated PIN or passcode, and that the head-movement signature is used for secondary verification. Upon entry of the correct password, the device verifies the identity of the user based on the head-movement signature.

In the head-movement authentication phase, given a test sample, we classify the sample as true (1) or false (0) based on one of the two classifiers discussed above. If the result is true, the user is accepted; otherwise the user is rejected.

4. EVALUATING HEADBANGER SYSTEM

We conducted extensive evaluation of the Headbanger system. In particular, our evaluation focused on showing that a user's body movement pattern can be used to establish the user's identify. We first show that even a simple movement pattern is hard to imitate by others (i.e., being distinctive), and next show that people can easily repeat their own patterns (i.e., being repeatable). To conduct the evaluation, we prototyped the Headbanger system using the Google Glass, but our system can be easily implemented on other platforms.

4.1 Dataset

We collected head-movement ACC samples from 24 subjects. We informed the participants of the risks involved in wearing Google Glass and collecting head-movement data (e.g., feeling dizzy after head-movement for a period of time, not being able to see clearly if near-sighted, etc.). If they agreed to participate, we asked each subject to report their age and gender. We asked the subjects to wear the Google Glass and make head-movements while listening to the music track. Meanwhile, we recorded the raw accelerometer data. Each recording session was 10 seconds long, and we sat through each session with the subject to help him/her use the system properly. After every 5 recording sessions, we asked the subject to take a break to relax their muscle and regain energy. Both our system design and our data collection protocol were approved by our Institutional Review Board. The subjects, *** females and *** males, had an average age of *** years.

4.2 Even Simple Body Movement is Hard to Imitate

In the first set of experiments, we show that even simple body movement is very hard to imitate by others, and thus body movement is rather distinctive among different people.

4.2.1 Experimental Setup

One of the simplest body movement patterns that can be easily captured by built-in sensors is rhythmic nodding. ***YZ: can we show pictures of simple nodding?? *** Figure **??** shows the nodding pattern we employed in this set of experiments. The audio file we used in the experiments can be downloaded at ***.

Each ACC sample was collected for the duration of 10 seconds. When calculating the classification results, we generated shorter samples of varying lengths (2 seconds, 3 seconds, 6 seconds, and 10 seconds) from the original 10-second samples. The sum of the sample duration and the subsequent processing latency thus determines the authentication response time of our system. Longer sample durations likely lead to more accurate classification results, but shorter sample durations may be preferred by users due to faster authentication responses.

In this set of experiments, we have one glass owner, who designed the nodding pattern shown in Figure **??**, and 15 imitators who imitated the movement. We collected 100 10-second ACC samples from the owner, during the course of *** days (from *** to ***), ensuring the owner's sensor data includes sufficient variation that naturally occurs with time. We also made a great deal of effort to make sure the imitators accurately imitate the owner's movement – the owner carefully explained his movement pattern to each imitator, and sat through each data collection session for all the 15 imitators to make sure their movement pattern looks the same to the owner's eye. For reach imitator, we collected 40 10-second ACC samples.

Sample duration (s)	2			3		
	FRR (%)	FAR (%)	BAC (%)	FRR (%)	FAR (%)	BAC (%)
SVM Top 1	25.0	16.74	79.12	15.0	14.05	85.47
SVM Top 3 Voting	14.84	9.24	87.95	18.48	6.85	87.33
Sample duration (s)	6			10		
	FRR (%)	FAR (%)	BAC (%)	FRR (%)	FAR (%)	BAC (%)
SVM Top 1	3.33	6.66	95.0	0.0	9.62	95.18
SVM Top 3 Voting	7.27	6.17	93.27	4.84	6.17	94.48

Table 1: Average FAR, FRR, and BAC for SVM-based classification when we choose different 4 imitators in the training set (from the total 15 imitators). We have the results for different sample durations. In these results, we use the earliest 40 owner samples in the training set.

4.2.2 *Classification Metrics*

In this study, we consider both SVM based classification and thresholding based classification. The classification process is different for these two approaches:

- *SVM based classification.* For SVM based classification, we need to construct a training set that consists of both true training data (from the owner) and false training data (from the imitators). In total, we have 100 data samples from the owner and 15×40 samples from imitators, from which we include S samples from the owner, and all the samples from I imitators ($I \times 40$ samples) in the training set. Therefore, we have $100 - S$ owner samples, and $(15 - I) \times 40$ imitator samples in the testing set.
- *Thresholding based classification.* For thresholding based classification, the training set only consists of data samples from the owner. Similar to the SVM case, we use S samples from the owner as the training set, and all the other samples as the testing set.

In our evaluation, the default value for S is 40, and the default value for I is 4. In the evaluation, we varied the value of S and I as well as the chosen training samples to ensure the robustness of our system [2].

For both SVM-based and thresholding-based classification approaches, we consider two testing methods. The first method involves comparing the testing sample against the top 1 training sample from a user, which we refer to as *Top 1* testing. The second method involves comparing the testing sample against the top 3 training samples from a user and then determining the classification result by voting among these three results, which we refer to as *Top 3 Voting* testing.

In this study, we consider classification metrics that are popular in biometric authentication, namely false acceptance rate FAR (the percentage of false testing samples that are mistakenly accepted) and false rejection rate FRR (the percentage of true testing samples that are mistakenly rejected). Usually, an overly strict classification system leads to high FRR, while an overly relaxed system leads to high FAR. We also consider balanced accuracy

$$BAC = 1 - (FAR + FRR)/2,$$

which is a combined metric that measures both FAR and FRR.

4.2.3 *SVM Classification Results*

As mentioned earlier, SVM classification requires training data from both the owner and imitators. In the results shown in Table 1, we used the earliest 40 samples from the owner (out of 100 total owner samples), but varied the imitators that are included in the training set. Specifically, if we label the 15 imitators as $I_1, ..., I_{15}$, then we had the following 15 different combinations of imitators in the training set: $\{I_1, I_2, I_3, I_4\}$, $\{I_2, I_3, I_4, I_5\}$, $\dots$, $\{I_{15}, I_1, I_2, I_3\}$. Table 1 shows the average FAR, FRR, and BAC values over these 15 combinations. By considering these different cases, we can eliminate the impact of having a specific imitator in the training data and the results are more representative. Here, we also vary the sample duration: 2, 3, 6, and 10 seconds.

From Table 1, we have the following main observations:

1. *Even simple nodding is not easy to imitate: nodding for 6 seconds can classify 95% of the users.* As the sample duration increases from 2 seconds to 6 seconds, the classification accuracy improves significantly – the BAC value changes from 79.12% to 95% for top 1 testing, and from 87.95% to 93.27% for top 3 voting. After the sample duration reaches 6 seconds, the improvement becomes less pronounced. This suggests that a sample duration of 6 seconds is sufficient to successfully classify 95% of the users. We feel that moving our head gently along with music for 6 seconds is in general not a cumbersome process for most users. It further suggests that even simple nodding is hard to imitate by others, and thus head-movement has the potential to serve as a reliable biometric characteristic for smart wearable authentication.

2. *For sample duration of 6 seconds, testing against 1 training sample is sufficient.* The results also show that, when the sample duration is short (e.g., 2 or 3 seconds), comparing the testing sample against top 3 training samples leads to much better classification result. When the sample duration is reasonably long, say 6 seconds, comparing the testing sample against the top 1 training sample leads to better results. This is advantageous because top 1 testing incurs much less processing and energy overhead, lending itself for execution on wearable devices.

The Impact of Training Dataset Size: Next we studied whether the number of owner samples in the training set has a bearing on the SVM classification results. We varied the number of owner training samples as 10, 20, and 30, and show the resulting FRR and FAR values in Figure 6(a) and BAC results in Figure 6(b). Note that we have in total 100 samples from the owner. ***YZ: 40??***

From the results, we observe that there is no clear trend when we vary the number of owner samples in the training set. When we increase the owner training samples from 10 to 20, the average BAC value increased, but only very marginally. This result suggests that to make the classification light-weight, thus suitable for wearable devices, we can use a small number of training set without hurting the performance.

[2] In the interest of space, we didn't include the results with different I values in this paper.

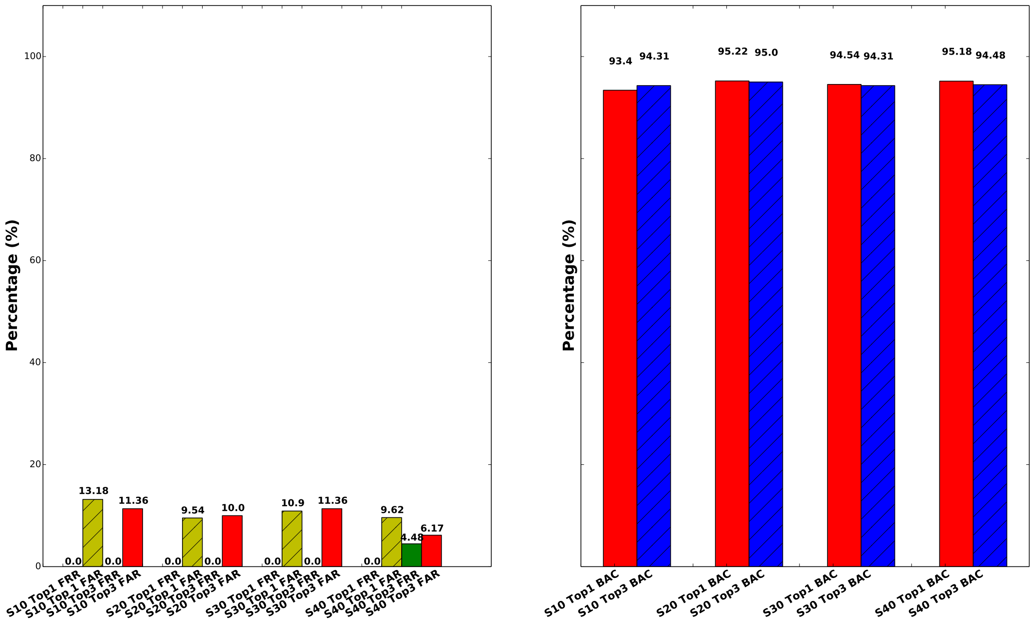


Figure 6: In this set of experiments, we studied whether the number of owner samples in the training set has a bearing on the SVM classification results. (a) shows the FRR and FAR results for each scenario, and (b) shows the BAC results.

4.2.4 *Thresholding-Based Classification*

Thresholding based classification only requires owner samples in the training set. In order to eliminate the impact of choosing certain training samples, we rotate the 40 training samples and summarize the average FAR, FRR, and BAC values in Table 2. Specifically, if we label the 100 owner samples as $S_1, S_2, ..., S_{100}$, then we ran 100 different experiments in this case, with training set being $\{S_1, S_2, ..., S_{40}\}$, $\{S_2, S_3, ..., S_{41}\}, \ldots, \{S_{100}, S_1, ...S_{39}\}$, respectively. Here, we also varied the sample duration: 2, 3, 6, and 10 seconds.

From Table 2, we have the following main observations:

1. *Thresholding-based classification, though more lightweight, fares better than SVM.* Results in Table 2 show that thresholding-based classification in general fares better than SVM (except the 6-second case). With 10-second sample duration, thresholding-based classification has a BAC value of 98.33%. This is mainly because the thresholding-based classifier only relies on true training samples, hence more robust against imitators whose movement pattern is similar to that of the owner. This result is desirable because thresholding-based classification is much more light-weight than SVM classification, and thus more suitable for wearable devices.

2. *For sample duration of 6 seconds or longer, comparing against the top 1 sample is sufficient.* We also observe that when the sample duration is 6 seconds or longer, it is preferred to compare the test sample against the top 1 training sample. This can significantly reduce the computation overhead of our system.

3. *Thresholding-based classification has better FRR than FAR values.* Results in Table 2 show that thresholding-based classification always has very low FRR values, even with short sample durations. With sample duration of 2 seconds, the FRR values are 3.33% for top 1 testing and 1.80% for top 3 voting, while the FAR values are 18.57% for top 1 testing and 13.94 for top 3 voting. This suggests that even with short sample durations, thresholding-based classification is good at identifying the pattern of the same user. When the sample duration increases, we observe that the FAR values drop significantly, to as low as 1.99% for top 1 testing and 4.78% for top 3 voting for 10-second sample durations.

The Impact of Training Dataset Size: Next we studied whether the number of owner samples in the training set has a bearing on the thresholding classification results. We varied the number of owner training samples as 10, 20, and 30, and show the resulting FRR and FAR values in Figure 7(a) and BAC results in Figure 7(b). ***YZ: 40??***

Like the case in SVM classification, we didn't see any noticeable change in classification results with respect to the training data set size. This observation is advantageous in that we can reduce the training dataset size to further reduce the processing requirement and power consumption of our system.

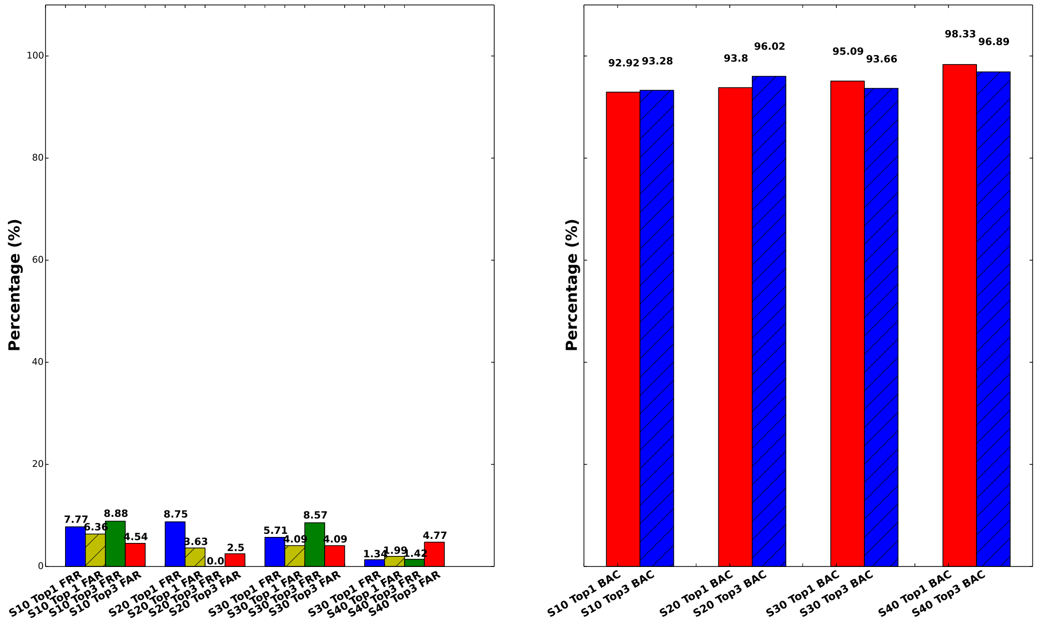


Figure 7: ***YZ: need to change this plot* In this set of experiments, we studied whether the number of owner samples in the training set has a bearing on the thresholding classification results. (a) shows the FRR and FAR results for each scenario, and (b) shows the BAC results.**

4.3 Head-movement Patterns are Repeatable

In the second set of experiments, we would like to find out whether a user can successfully repeat her own head movement if each user is asked to come up with their own movement pattern.

Classification Results: In this set of experiments, we had 8 subjects, and for each subject, we collected 38 ACC samples with sample duration of 10 seconds. Each subject performed different head-movement patterns of their choice.

In generating the classification results, we employ a process that involves 8 iterations. In each iteration, we assume the legitimate glass owner is one of the users. We use 19 ACC samples of this user as the training data set, and testing with his/her other 19 ACC samples together with the other 7 user's ACC samples. For each iteration, we calculate the resulting FRR, FAR, and BAC values (using the top 3 voting for thresholding-based classification), and show the results in Figure 8, where Figure 8(a) shows the FRR and FAR values, while Figure 8(b) shows the BAC values.

After carefully examining the results in Figure 8, we have the following main observations:

1. *Head-movements are highly repeatable.* The first observation is that a user's head-movement pattern is usually highly repeatable. Among the 8 subjects that we studied, the highest BAC value is 100%, and the lowest is 91.81%, with the average BAC value of ***%. Overall, this result suggests that the head-movement pattern is a promising biometric candidate for user authentication.
2. *User head-movements have different level of stability.* Even from the small-scale 8 subject study, we found that users have different level of stability in their head-movements; some are much more stable than others. In our example, subject 3 has 0 FRR and 0 FAR, while subject 1 has 5.26% FRR and 11.11% FAR. We suspect that through proper training, subjects could improve their stability. However, we didn't test this hypothesis in this study. We also didn't notice any correlation between a user's stability with his/her age or gender.

4.4 Headbanger Authentication App Implementation

We implemented the Headbanger authentication app on Google glass. ***YZ: maybe we can have a software block diagram for the app, and then describe where each block is executed. We should also have some pictures.***

5. DISCUSSION AND FUTURE WORK

In this study, we showed that head-movements have the potential to be used as a reliable biometric characteristic for user authentication. We will now discuss some of the limitations that we identified from this work and prospects for future work as below.

5.1 Reliability

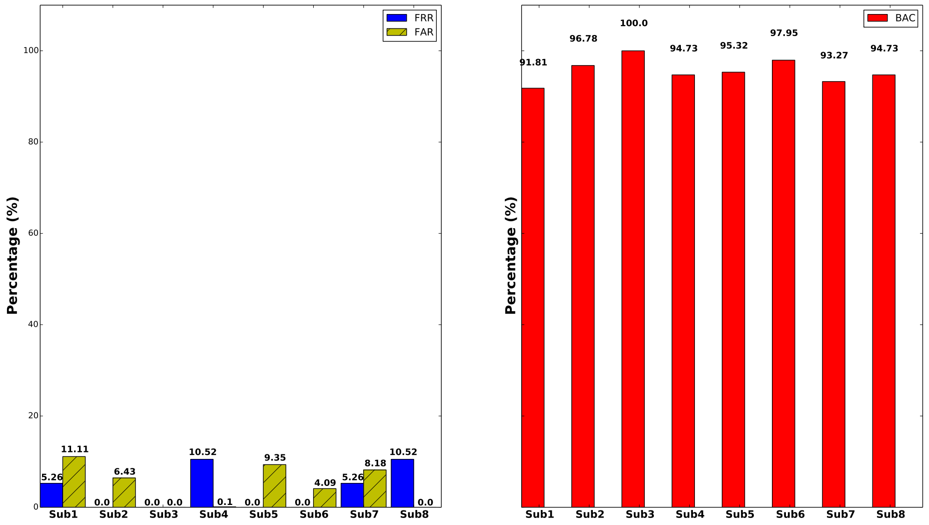


Figure 8: In this set of experiments, we studied whether a user can successfully repeat her own head-movement pattern. We had 8 subjects, each performing her own choice of head-movement patterns. We collected 38 samples for each subject. (a) shows the FRR and FAR results for each subject, and (b) shows the BAC results. Thresholding-based classification with top 3 voting was used to generate these results.

Our work in this paper shows that head-movements are distinctive and repeatable in controlled settings. However, in reality, the behavior of head-movement signatures over chaotic settings will be a key factor to decide on the effectiveness of this approach. Our work only evaluates the case when a user is in a stationary setting when attempting to login, such as when sitting on the chair or standing still. The performance of this approach in realistic mobile settings such as walking or in a vehicle is yet unknown. It is also unclear if the head-movement patterns are repeatable in such a mobile environment, or if the ambiguities of vehicle motion versus head-motion can be separated. Similarly, a person's head-movement signature may also depend on the mood/energy level of the person; for example, a fresh and energetic user may provide significant head-movements as compared to a sick or tired user whose signatures may not even be detectable. Inconsistencies in the accelerometer sensor such as drift and temporal bias can significantly affect the nature of inferred head-movement signature. Head-movements, on the other hand, may also evolve over time for a person which call for periodic calibration of the system and/or the training data.

While reliability metric was out of the main scope of this paper, we are keen to address the same in future work.

5.2 Multi-Modality

Smart-glass devices typically contain an array of motion sensors such as accelerometer, gyroscope, IMU. It is only a matter of time that motion sensor chips will be integrated into wearable devices. This opens up opportunities for multi-modal motion sensing. For example, accelerometer data can be combined with gyroscope measurements to provide multi-dimensional head-movement features that can improve the quality of the inferred signatures. Head movements can also be combined with other body movements to generate valuable, reliable signatures for authentication. For example, through a simple test experiment using the Google Glass infra-red light sensor[3] we observed that the blinking and winking patterns of users in response to the music stimulus were reasonably differentiable among users. Such patterns may also independently serve as another biometric that can be used for authentication purpose, or can be combined with head-movements for better results. Recent studies have shown that heart beat or pulse can also serve as reliable biometric for authentication purposes [17, 1] We reserve such potential enhancements to our system for future implementation.

5.3 Seamless Protocol

An authentication system must have an effective protocol for authenticating users seamlessly to their device. Our system runs a simple authentication protocol where the user is given a finite set of (calibrated) music tracks to pick, and based on which the user makes head-movements in response. Our design assumes that the user voluntarily accepts the enforcement of the requirement of head-movements in response

[3] we had to root the Google Glass to access the IR sensor unit.

to the music. A seamless design would ensure that the system captures even the slightest of the subconscious head-movements in the event that the user does not make any enforced head-movements. In such cases the head-movement signatures will have to be much more elaborate with multiple attributes that correspond to the different realistic use-cases of the system.

5.4 Processing/Battery Power Constraints

Battery power consumption and computing power are very important parameters for consideration when optimizing a design to accommodate to wearable devices. Wearables usually have serious resource limitations, especially in terms of processing power and battery power. This paper, addresses these concerns through optimization strategies in the head-movement signature classification stage of the proposed authentication algorithm. For wearables it is important that such optimization strategies are taken to next levels until a roadblock is reached. For example, one such strategy extension in this work can be that, after a short duration, before the entire music is played, if it is found that a user's movement does not match the signature of the claimed user with sufficient confidence level, then the on-site classification may be terminated instead of waiting for the entire duration to yield the rejection. Another example, may include cyber-foraging strategies to offload heavy computation tasks, such as classification, to the user's Bluetooth paired smartphone.

5.5 Large-scale evaluation

To be adopted as a primary authentication mechanism on smart-glass devices, the technique will have to be evaluated over a large number of usage and and over a large user base. Conducting such rigorous large-scale experiments are typically infeasible in academic laboratory settings. We reserve such large scale experiments for future work, and hope to accomplish through industry collaborations.

6. RELATED WORK

Broadly speaking, Headbanger is a 3D behavioral biometric based activity recognition system. More specifically, it does the job of head gesture based authentication. To our best knowledge, it is the first wearable authentication system that only relies on accelerometers signature data from human *head* motion. In this section, we briefly review the related literature on mobile device authentication.

Sample duration (s)	2			3		
	FRR (%)	FAR (%)	BAC (%)	FRR (%)	FAR (%)	BAC (%)
Th Top 1	3.33	18.57	89.04	5.76	16.63	88.8
Th Top 3 Voting	1.80	13.94	92.12	1.04	12.67	93.14

Sample duration (s)	6			10		
	FRR (%)	FAR (%)	BAC (%)	FRR (%)	FAR (%)	BAC (%)
Th Top 1	3.17	8.74	94.05	1.34	1.99	98.33
Th Top 3 Voting	5.63	9.15	92.60	1.42	4.78	96.89

Table 2: Average FAR, FRR, and BAC for thresholding-based classification when we choose different 40 owner samples in the training set (from the total of 100 samples). We have the results for different sample durations. In these results, we use the first four imitators in the training set.

Harwin et al. [16] was considered the first one proposed to use head gestures by combining pointing and movements for human computer interaction. The closest work to ours is Prescott [31], which used with human blinking pattern with the song as stimuli as a unique feature for authentication. They achieved 82.02% accuracy among 9 individuals based on how they blink based on the same pattern. We share the same philosophy that users will generate their unique biometric signature with the song assistance, but use head motion instead for authentication. Ishimaru et al. [18] proposed to k combine the eye blink frequency from the infrared proximity sensor and head motion patterns from accelerometer sensor on Google Glass to recognize different activities (reading, talking, watching TV, math problem solving and sawing), and achieved 82% recognition accuracy. Our approaches differ from theirs that we only use head motion and we focus on the problem of human verification under the same activity activity instead. There are also a number of head motion based activity recognition work using computer vision, such as [22]. BioGlass [17] combines Google Glass's accelerometers, gyroscope, and camera to extract measure physiological signals of the wearer such as pulse and respiratory rates.

Other than head, accelerometers have been placed in other parts of body for gait-based authentication purpose, such as waist [3], pocket [14], arm [26, 15], leg [13] and ankle [12]. They share the same thread that collect the raw accelerometers data and apply signal processing and/or machine learning techniques on top of the data dynamics induced by human motion to perform authentication.

Hand gesture and touchscreen are often used naturally in pair for authenticating that device. A number of contexts information including biometric features (e.g. finger length, hand size, swipe/zoom speed, acceleration, click gap, contact size, pressure) and behavioral feature (e.g. touch location, swipe/zoom length, swipe/zoom curvature, time, duration) have been exploited as effective features for authentication purpose such as demonstrated in [28, 11, 7, 10]. While most of the techniques require users to explicitly conduct a gesture following a specific pattern, TIPS [10] proposed a Multi-Stage Filtering with Dynamic Template Adaptation strategy to perform the user authentication in a uncontrolled environments – as user naturally use the phone.

Finally, we acknowledge that there are a number of other authentication using other techniques, such as speech [27], computer vision and image [6], graphical password [4, 29], biometric fingerprints [19]. Due to the space constraints and their indirect relevance to our work, we do not list them here. However, our approach can be used as a complementary scheme to most existing technique, which leaves space new scheme of multimodal approach design for future work.

7. CONCLUDING REMARKS

We developed a system that uses head-movement patterns of users for direct authentication to a wearable device. We developed a novel approach that infers head-movements of users in response to music and generates signatures that are unique to every user. Experimental observations revealed that the head-movement signatures generated using the dynamic time-warping tool, in response to the same music track, were consistent in typical stationary environments such as; user sitting or standing at one location while attempting to authenticate. We leveraged the consistency in the head-movement signatures to develop two classification algorithms, based on machine learning and adaptive thresholding, to efficiently and accurately label user signatures. Through a multiple user based experiment evaluation using the data collected by our sensor data collection app in Google Glass, we observed that the average true-acceptance rate of our approach is above ***% and the false detection rates is below ***%. The multi-user training data sets were validated and verified during the course of our evaluations and will be released for public use in near future.

8. REFERENCES

[1]

[2] H. Ailisto, E. Vildjiounaite, M. Lindholm, S.-M. Mäkelä, and J. Peltola. Soft biometricsâĂŤcombining body weight and fat measurements with fingerprint biometrics. *Pattern Recognition Letters*, 2006.

[3] H. J. Ailisto, M. Lindholm, J. Mantyjarvi, E. Vildjiounaite, and S.-M. Makela. Identifying people from gait pattern with accelerometers. In *Defense and Security*. International Society for Optics and Photonics, 2005.

[4] R. Biddle, S. Chiasson, and P. C. Van Oorschot. Graphical passwords: Learning from the first twelve years. *ACM Computing Surveys*, 2012.

[5] C. Bo, L. Zhang, X.-Y. Li, Q. Huang, and Y. Wang. Silentsense: silent user identification via touch and movement behavioral biometrics. In *ACM MobiCom*, 2013.

[6] K. W. Bowyer, K. Chang, and P. Flynn. A survey of approaches and challenges in 3d and multi-modal 3d+ 2d face recognition. *Computer vision and image understanding*, 2006.

[7] Z. Cai, C. Shen, M. Wang, Y. Song, and J. Wang. Mobile authentication through touch-behavior features. In *Biometric Recognition*. Springer, 2013.

[8] R. T. Collins, R. Gross, and J. Shi. Silhouette-based human identification from body shape and gait. In *IEEE FGR*, 2002.

[9] A. De Luca, A. Hang, F. Brudy, C. Lindner, and H. Hussmann. Touch me once and i know it's you!: implicit authentication based on touch screen patterns. In *ACM CHI*, 2012.

[10] T. Feng, J. Yang, Z. Yan, E. M. Tapia, and W. Shi. Tips: context-aware implicit user identification using touch screen in uncontrolled environments. In *ACM HotMobile*, 2014.

[11] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song. Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *IEEE Transactions on Information Forensics and Security*, 2013.

[12] D. Gafurov, P. Bours, and E. Snekkenes. User authentication based on foot motion. *Signal, Image and Video Processing*, 2011.

[13] D. Gafurov, K. Helkala, and T. Søndrol. Biometric gait authentication using accelerometer sensor. *Journal of computers*, 2006.

[14] D. Gafurov, E. Snekkenes, and P. Bours. Gait authentication and identification using wearable accelerometer sensor. In *IEEE AIAT*, 2007.

[15] D. Gafurov and E. Snekkkenes. Arm swing as a weak biometric for unobtrusive user authentication. In *IEEE IIHMSP*, 2008.

[16] W. Harwin and R. Jackson. Analysis of intentional head gestures to assist computer access by physically disabled people. *Journal of biomedical engineering*, 1990.

[17] J. Hernandez, Y. Li, J. M. Rehg, and R. W. Picard. Bioglass: Physiological parameter estimation using a head-mounted wearable device. In *IEEE MobiHealth*, 2014.

[18] S. Ishimaru, K. Kunze, K. Kise, J. Weppner, A. Dengel, P. Lukowicz, and A. Bulling. In the blink of an eye: combining head motion and eye blink frequency for activity recognition with google glass. In *ACM AH*, 2014.

[19] A. K. Jain, L. Hong, S. Pankanti, and R. Bolle. An identity-authentication system using fingerprints. *Proceedings of the IEEE*, 1997.

[20] A. K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 2004.

[21] Z. Jorgensen and T. Yu. On mouse dynamics as a behavioral biometric for authentication. In *ASIACCS*, 2011.

[22] R. Kjeldsen. Head gestures for computer control. In *IEEE ICCV Workshop*, 2001.

[23] R. E. Milliman. Using background music to affect the behavior of supermarket shoppers. *The Journal of Marketing*, 1982.

[24] F. Monrose and A. D. Rubin. Keystroke dynamics as a biometric for authentication. *Future Generation computer systems*, 2000.

[25] L. O'Gorman. Comparing passwords, tokens, and biometrics for user authentication. *Proceedings of the IEEE*, 2003.

[26] F. Okumura, A. Kubota, Y. Hatori, K. Matsuo, M. Hashimoto, and A. Koike. A study on biometric authentication based on arm sweep action with acceleration sensor. In *IEEE ISPACS*, 2006.

[27] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn. Speaker verification using adapted gaussian mixture models. *Digital signal processing*, 2000.

[28] N. Sae-Bae, K. Ahmed, K. Isbister, and N. Memon. Biometric-rich gestures: a novel approach to authentication on multi-touch devices. In *ACM CHI*, 2012.

[29] M. Sherman, G. Clark, Y. Yang, S. Sugrim, A. Modig, J. Lindqvist, A. Oulasvirta, and T. Roos. User-generated free-form gestures for authentication: Security and memorability. In *ACM MobiSys*, 2014.

[30] S. V. Stevenage, M. S. Nixon, and K. Vince. Visual analysis of gait as a cue to identity. *Applied cognitive psychology*, 1999.

[31] T. Westeyn and T. Starner. Recognizing song-based blink patterns: Applications for restricted and universal access. In *IEEE FGR*, 2004.

[32] R. V. Yampolskiy. Motor-skill based biometrics. In *Annual Security Conference*, 2007.

[33] M. Zentner and T. Eerola. Rhythmic engagement with music in infancy. *Proceedings of the National Academy of Sciences*, 2010.