Whose Move is it Anyway? Authenticating Smart Wearable Devices Using Unique Head Movement Patterns

Sugang Li*, Ashwin Ashok†, Yanyong Zhang*, Chenren Xu‡, Janne Lindqvist*, Macro Gruteser*
*WINLAB, Rutgers University, North Brunswick, NJ, USA
†Carnegie Mellon University, Pittsburgh, PA, USA
‡CECA, Peking University, Beijing, China

***Abstract*—In this paper, we present the design, implementation and evaluation of a user authentication system, *Headbanger*, for smart head-worn devices, through monitoring the user's unique head-movement patterns in response to an external audio stimulus. Compared to today's solutions, which primarily rely on indirect authentication mechanisms via the user's smartphone, thus cumbersome and susceptible to adversary intrusions, the proposed head-movement based authentication provides an accurate, robust, light-weight and convenient solution.**

Through extensive experimental evaluation with 95 participants, we show that our mechanism can accurately authenticate users with an average true acceptance rate of 95.57% while keeping the average false acceptance rate of 4.43%. We also show that even simple head-movement patterns are robust against imitation attacks. Finally, we demonstrate our authentication algorithm is rather light-weight: the overall processing latency on Google Glass is around 1.9 seconds.

I. Introduction

Wearable devices are on the way to become an integral part of people's daily lives [10], [16], [35]. These devices collect data about the wearers, their surroundings and often even about their health. It is thus critical to the users' privacy, that this data is protected from unauthorized access. Although there has been work [19], [20], [23] on limiting privacy threats from ubiquitous photography enabled by the wearable devices, robust usable and secure authentication systems leveraging the devices have not emerged. An authentication system for these devices has to strike an appropriate balance with user convenience, especially since users are interacting with an increasing number of wearables.

Authentication Challenge. Today, authentication on most commercially available wearable devices [10], [35] relies on an indirect mechanism, where users can log in to their wearables through phones. This requires the wearable device to be registered and paired to the mobile device, and both devices to be carried by the user, which can be highly inconvenient in reality. The security of this approach is also in question as it increases the chance of hacking into both devices if either is lost or stolen. Some devices including Google Glass [16] and FitBit's health tracker [10] allow linking the device to online accounts instead of a mobile device for convenience, which, however, does not make it more secure. Indirect authentication remains a dominant paradigm for wearables despite these fundamental shortcomings because such devices are *seriously resource-constrained* in many aspects: battery power, computational and storage capabilities, and input/output. As a result, typical authentication methods designed for more resource-rich devices can not be directly applied here; rather, user authentication for wearable devices must operate indirectly through a more capable device. We, however, take the viewpoint that wearables will become more independent units that have to maintain security guarantees without such paired devices and we seek to develop suitable *direct authentication* methods that are both accurate and light-weight.

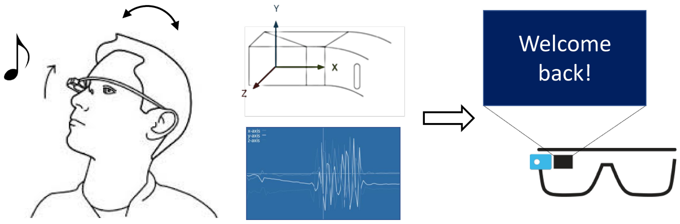


Fig. 1. Illustration of Headbanger. The head-worn device authenticates the users based on signatures generated from head-movement patterns. These patterns are created in response to an audio snapshot played on the device.

Before we explore direct authentication methods for wearable devices, let us first consider available solutions for other mobile systems, especially smartphones and tablets. Broadly speaking, the two most commonly used authentication methods on mobile systems are (arguably) password-based methods (with their variants) and biometric-based methods. However, we argue that neither of these two methods is really suitable for wearable devices. Typing passwords or drawing swipe patterns on wearable devices can be rather cumbersome due to their small input/output units, if they do have a touch sensor at all. Collecting and recognizing physiological biometrics (such as DNA, fingerprint, hand/finger geometry, iris, odor, palm-print, retinal scan, voice) requires specialized sensing hardware and processing resources that add cost, and many of these sensors are even larger than the size of wearables themselves.

We therefore focus on a third class of direct authentication methods: relying upon the uniqueness of human behavior characteristics such as human walking gait, arm swings, typing patterns, body pulse beats, eye-blinks, etc. This way of authenticating users is often referred to as behavior-based authentication, and it has been studied in the context of authenticating smartphones and tablets [6]–[8], [25], [28]–[30], [36]. The main advantage of using behavioral characteristics for mobile devices is that the signatures can be readily generated from raw data of built in sensors such as motion sensors, camera, and microphones. Considering that cameras and microphones, as well as vision and audio processing algorithms, are quite energy-hungry, we thus focus on those behavioral characteristics that can be easily captured by sensors that require less power consumption, such as accelerometer. More specifically, we propose to authenticate wearable devices to users based on the following behavioral characteristic: our unique body movement patterns and their dependence on external stimuli that wearable devices can generate, such as vibrations and music.

Head-movement based authentication. Body movement patterns have long been used by humans to discriminate between people. By watching how a person walks, dances, waves hands, we can often recognize the person from afar. This is because human body movements are usually *distinctive* and *repeatable*. Achieving the same through wearables, however, is not straightforward and poses significant research challenges: it is unclear whether these seriously-constrained devices are able to capture the differentiating characteristics of movement patterns, process the data, and quantify the uniqueness of each user's behaviors. Moreover, each device will have only a limited view of body movements, dependent on its mounting position on the human body. In this paper, we set out to conduct a holistic study of wearable authentication through body movements and to design an accurate, robust and lightweight authentication system. A key distinguishing feature of our work is that we will also consider stimuli that wearable devices can provide, particularly stimuli that are difficult to observe even for the closest adversaries. For example, we can use fast-tempo music through earbuds to stimulate movements and to make such free-style movements more repeatable.

In particular, we have designed, implemented and evaluated *Headbanger*, an authentication system that can authenticate users by sensing head movements when listening to music beats. Although we use Google Glass as a running example, 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, augmented reality devices.

In summary, the key contributions of this paper are:

1) We have designed and implemented a novel user authentication method for wearable devices using head-movement patterns. Our study shows that user's head-movement patterns contain unique signatures that when inferred correctly can be used as valid means for authentication. We design a system, *Headbanger*, that records, processes, and classifies head-movement patterns of users based on the built-in accelerometer sensor readings.
2) Through comprehensive experiments involving 95 participants and over different system design parameters we show that head-movement patterns can generate accurate authentication results. Our approach effectively identifies a wearable device user, with an average false acceptance rate of 4.43% and an average true-positive rate of 95.57%. Also, we show that our authentication method is quite robust: when a user slightly increases her head-movement complexity, it quickly becomes *much harder* for attackers to imitate the movement pattern.
3) We implement *Headbanger* on Google Glass and carefully profile the execution time of each software module in the implementation. Our measurements indicate an average processing latency of 1.93 seconds on the Google Glass for the best authentication performance.

II. Background

A. Mobile Device Authentication Through Body Movements

A number of body-movement based authentication approaches have been proposed for mobile devices. These systems leverage unique signatures from human behavior that may be subconscious or in response to external stimulus or both. 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 [36]. Okumura et. al. [29] have shown that human arm swing patterns can be used to create signatures to authenticate to their cell-phones. Monrose et.al. [28] 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 to authenticate users. Similarly, mouse usage dynamics [25] and touchpad touching dynamics [3], [7] have also been shown to serve as potential authentication cues.

We take the viewpoint that, in comparison to other means of authentication, body-movement based authentication may offer great convenience. With this rationale, 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.

B. Using Head-movement for Authentication

According to Jain et al. [22], a type of body movement is useful for authentication when it is *universal*, *distinctive*, *repeatable*, and *collectible*. Sensors for collecting head-movement patterns are available on most of today's head-worn wearable devices, and thus making head movements both *universal* and *collectible*.

In this paper, we show that free-style head movements are *distinctive* and *repeatable*, especially when combined with external stimuli such as music beats. In *Headbanger*, music plays a crucial role in stimulating body movements such that the resulting movement pattern is natural to the user (more distinctive) and easier to remember (more repeatable). Zentner and Eerola [38] have shown that most people move their body

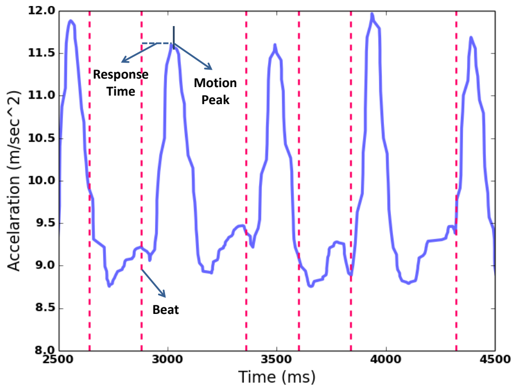


Fig. 2. The response time of a head motion is the interval between the motion and the music beat to which the motion responds. From a sequence of head motions, we can obtain the response time sequence.

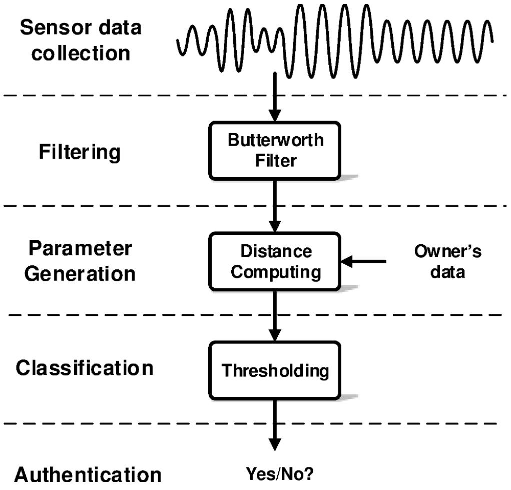


Fig. 4. *Headbanger* system design flow. The online authentication phase of *Headbanger* consists of the following steps: (1) sensor data collection in which we collect accelerometer data while users move their head as a response to an audio track played on the glass, (2) filtering in which we apply a Butterworth filtering to smoothen the sensor data for subsequent processing, (3) parameter generation in which we calculate the distances between two accelerometer samples as the parameter, and (4) classification in which we adopt an adaptive thresholding mechanism to classify the user's head movement, whose output will be used as the authentication result.

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. Most adults naturally perform head movements or hand movements when listening to a fast beat audio track [24]. When combined with external rhythmic stimuli, we believe body movements become more distinctive – not only a person's movement pattern is unique, but their response to rhythmic stimuli is also unique. In this way, the resulting authentication system will be more dependable.

C. Motivation for Headbanger

Next, we conducted a preliminary experiment to investigate whether head-movement patterns can be potentially used to authenticate users. In this experiment, we collected head-movement accelerometer data from 28 subjects, wherein each subject was asked to perform a simple nodding movement following a short audio track (referred to as *music cue* in the rest of the paper). For this purpose, we examined a simple aspect of the head-movement pattern, response time, which is the time interval between a music beat and the corresponding nodding motion, as shown in Figure 2. The response time indicates how quickly a person responds to music beats.

In this experiment, we collected 20 samples from each subject, hence 20 response time time series for each subject. Then we calculated the average distance scores between these response times, both for the same subject and across subjects. We considered three types of distance metrics: cosine distance (COS), correlation distance (COR), and dynamic time warping (DTW) distance [2], and the readers can find the detailed description of these three metrics in Section III-B. We plot these three types of distance values in Figures 3 (a)-(c), respectively. In each plot, we include distance values for all 28 subjects – for each subject, we plot average distance scores between her and each of the other 27 subjects (referred to as distances with false subjects, shown in blue dots), as well as average distance scores among her own response times (referred to as distances with true subjects, shown in red squares). All three figures clearly demonstrate that the average distance score between a subject's samples is much lower than that among different subjects' samples, which further suggests that a subject exhibits repeatable and unique head nodding characteristics.

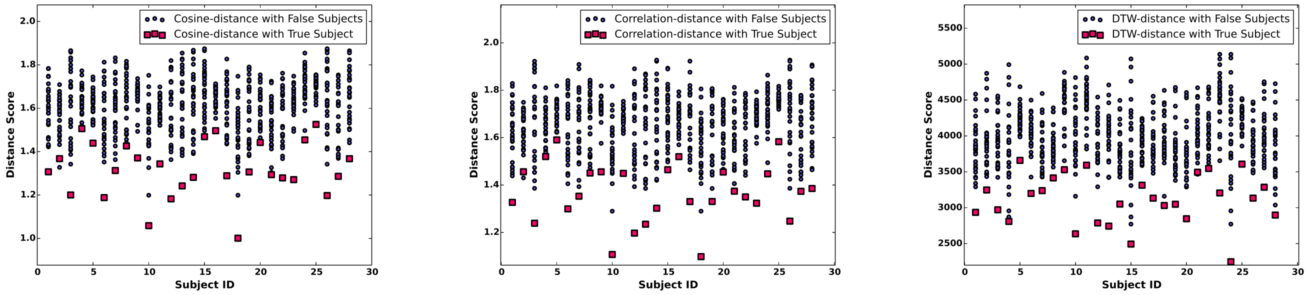
These observations suggest that even with simple nodding movements, the accelerometer data collected by Google Glass have the potential to be used for accurate and robust authentication. Motivated by this observation, we next devise the *Headbanger* authentication system.

III. Headbanger System Design

Headbanger enables direct authentication of users to their smart-glass devices or smart-glass apps using head-movements. We posit that *Headbanger* will run as a service in the device upon power-up or application start-up, similar to the screen-lock in smartphones.

The authentication process has two phases: an offline training phase and an online authentication phase. In the training phase, the system collects sensor readings when the real user moves her head with a music cue (following her pre-designed movement pattern), and uses the collected data to build a classifier. In the following discussion, we assume there is only one real user for the device for the sake of simplicity. An extension to support multiple users per device can be realized through minor modifications, namely, by appropriately indexing the users in the trained database.

In the online authentication phase, we collect sensor samples during a user's authentication attempt and label the samples using the *Headbanger* classifier. The user is authenticated upon a successful classification.



(a) Cosine Distance

(b) Correlation

(c) DTW

Fig. 3. (a) Cosine, (b) Correlation and (c) DTW distances are computed over 20 response time time series for each subject, with 28 subjects in total. For each subject's column, a red square represents the average distance among that subject's own response times, while a blue dot represents the average distance to other subjects' response times. The results show that red distances are always lower than blue ones, suggesting that different subjects' head-movement patterns are distinguishable.

As illustrated in Figure 4, *Headbanger* involves the following key modules: sensor data collection and filtering, sample distance computing, and classification. We will now discuss these design aspects in more detail.

A. Sensor Data Collection and Filtering

The data collection step involves collecting motion sensor data (we mainly focus on accelerometer in this study) while the user makes head-movements in response to the music cue with a duration of T seconds. The raw accelerometer signals are collected at a sampling rate of r samples/sec. The accelerometer data corresponding to one user, is a collection of accelerometer readings on the 3D axis (x, y, and z) collected over T-second duration, stored in a matrix with dimensionality $3 \times rT$. We will refer to this $3 \times rT$ matrix as a *sample*. 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.

Next, we filter the raw samples to remove noises due to spurious movements such as vibration or shaking. We adopt a low-pass digital Butterworth filter [4] and set a relaxed cut-off frequency of 10Hz.

B. Sample Distance Computing

In this study, we build a distance-based classifier for its simplicity is well suited for wearable devices. There are various ways of computing distances between two signals; we have considered three popular distance-computing algorithms in this study – Cosine (COS)distance, Correlation (COR)distance, and dynamic-time warping (DTW)distance.

Suppose we have two time series $S_a = (s_1, s_2, ..., s_n)$ and $S_b = (s_1, s_2, ..., s_n)$. Their COS distance is calculated as $\frac{S_a \cdot S_b}{\|S_a\| \times \|S_b\|}$; The COR distance is calculated by dividing their distance covariance by the product of their distance standard deviations; The DTW distance is defined as the distance when the optimal alignment between S_a and S_b has been determined by "warping" them non-linearly [2].

C. Classification

The classification step labels a test sample as "true" or "false" depending upon whether its distance to the real user's training samples is below a threshold. Again, we choose this method because it strikes a good balance between simplicity and performance. Next, we explain how we build the classifier and how to conduct online classification in detail:

1) *Identifying Top-K Training Samples.* Given M training samples, we first identify the K samples that are closest to all the training samples. For each training sample, we calculate its average distance to the other $M - 1$ samples, and then choose those K samples that have the lowest average distance values. These K samples are empirical estimation of the centroid of the sample space, and thus best represent the space among the collection of the training samples. We refer to them as Top-K samples. In our classifier, we focus on the Top-K training samples instead of all the training samples because it does not only incur much less computing overhead, but it also provides much better robustness against noises in training data.
2) *Establishing Distance Threshold.* Suppose a sample, s, is one of the Top-K samples. We have its distance scores to the other $M - 1$ samples in the training set, from which we can calculate the distance mean μ_s and distance standard deviation σ_s. Then sample s's distance threshold is defined as $(\mu_s + n\sigma_s)$, where n is referred to as the threshold parameter for our distance-based classifier.
3) *Classifying Test Sample.* If we use a training sample s to classify the test sample t, then t is labeled as a true sample if the distance between s and t is less then s's distance threshold $(\mu_s + n\sigma_s)$; otherwise, it is labelled as a false sample. The strictness of this classifier is characterized by the value of the threshold parameter, n; a large n can increase the false acceptance rate while a small n value can result in a high rejection rate of true samples.
4) *Voting*. We label the test sample according to all K Top-K samples, and the final classification result is the majority decision among all K individual classification results.

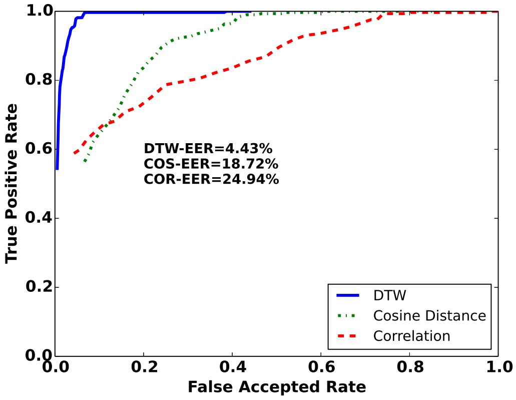


Fig. 5. The impact of the distance computing algorithm (i.e., DTW, cosine distance, and Correlation). In this set of results, we varied the value of n from 0.0 to 10.0 with an increment of 0.1, resulting in 100 thresholds in each case. We then plotted the TPR (y-axis) and FAR (x-axis) for each threshold. The results show that DTW delivers much better accuracies than the other two distance algorithms. As a result, in the remaining of the study, we will choose DTW for distance computing.

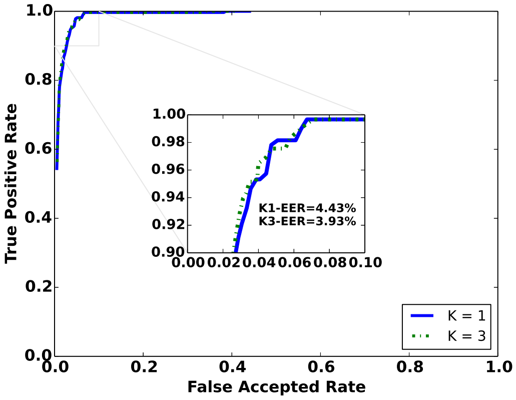


Fig. 6. The impact of different K values: $K = 1$ and $K = 3$. In this set of results, we varied the value of n from 0.0 to 10.0 with an increment of 0.1, resulting in 100 thresholds in each case. We then plotted the TPR (y-axis) and FAR (x-axis) for each threshold. The results show that voting schemes ($K = 3$) provide minor improvement on the performance. As a result, in the remaining of the study, we will choose $K = 1$.

Among the four steps outlined above, the first two steps belong to the offline training phase, while the last two steps belong to the online authentication phase. Finally, if the user's test sample is classified as "true" then the user is authenticated to the device; otherwise, the user is rejected.

IV. Evaluation

We conducted comprehensive evaluation of *Headbanger* through laboratory studies with human subjects – our studies were approved by the Institutional Review Board (IRB) of our institution. In the first phase of evaluation, we collected from volunteer participants accelerometer sensor readings with Google Glass. We analyzed these traces offline on a PC. Our evaluation in this phase is primarily aimed at determining the accuracy and robustness of *Headbanger*. In the second phase of evaluation, we implemented a *Headbanger* app and measured its processing latency. Our measurements suggest that *Headbanger* is indeed light-weight and can be executed on wearable devices such as Google Glass.

A. *Authentication Accuracy of* Headbanger

1) Participants: We had a total of 30 volunteer participants for this experiment, including a total of 19 males and 11 females. The mean age of the participants was 29.7 years with a standard deviation of 9.81 years. The youngest participant was 23 years old while the eldest was 54 years old.

2) Procedure: Our first experiment setup aimed at emulating the typical usage scenario of *Headbanger* for authentication, where a user conducts head-movements in response to a music cue played on the Google Glass device during a login attempt. In this experiment, all participants were asked to wear a Google Glass device. Participants who originally wore glasses (e.g. corrective eyewear) were asked to remove their glasses before conducting the experiment. The trials were conducted in an academic environment and overseen by one of our team members. The Google Glass ran our data-collection app that played a piece of music (music cue) for a specific duration, and recorded the accelerometer sensor readings. The sensor readings were recorded into a text file that was stored in the Glass's memory and later transported to a PC for offline processing through a Python script. The experiment was conducted in a well-lit indoor academic laboratory environment.

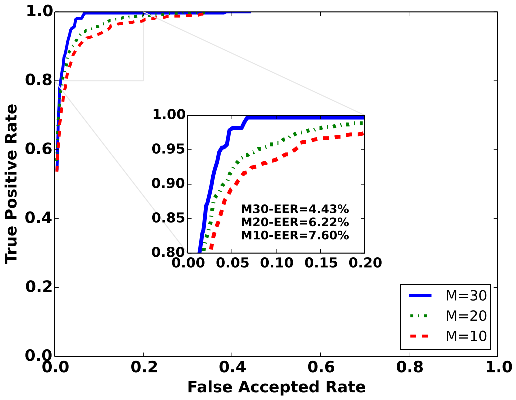


Fig. 7. The impact of training dataset size: 10, 20, and 30 samples. In this set of results, we varied the value of n from 0.0 to 10.0 with an increment of 0.1, resulting in 100 thresholds in each case. We then plotted the TPR (y-axis) and FAR (x-axis) for each threshold. The results show that having 30 samples delivers the best performance without adding to the online authentication computing overhead. As a result, in the remaining of the study, we will choose to have 30 samples.

During the course of a data collection session, the participants were allowed to take a break or withdraw from data collection if they felt uncomfortable at any point. The participants also could take a break for a minute after each trial. Each trial lasted for the duration of the music cue played

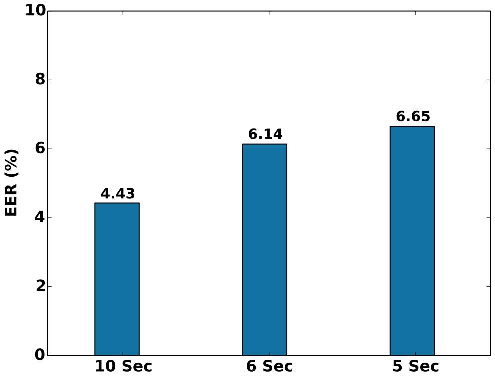


Fig. 8. The EER value of *Headbanger* when users choose different music cue lengths (10 sec, 6 sec and 5 sec). We have an EER value of 6.65% with a 5-second music cue, and 4.43% with a 10-second music cue, which is better than results of many similar body-movement based authentication systems (such as those in [13], [31]).

on the Glass, and a total of 40 such trials were conducted for each of the 30 participants. The entire data collection effort lasted over a duration of 60 days, of which 15 participants conducted their trials in a single sitting over a period of two hours, while the rest of the trails were spread over 3 days on average per subject.

3) Metrics: We evaluate the accuracy of *Headbanger* using metrics that are commonly used in evaluating authentication systems, namely, true positive rate TPR (percentage of true test samples that are correctly accepted), false acceptance rate FAR (percentage of false test samples that are mistakenly accepted), and true rejection rate TRR (percentage of true test samples that are mistakenly rejected). These three metrics are, however, largely dependent on the choice of the classification threshold parameter in *Headbanger*– a strict threshold in the classifier can lead to a high TRR value, while overly relaxing the threshold can lead to a high FAR. Hence, in order to report the threshold-independent performance, we also consider equal error rate EER which is the percentage of errors when we have $FAR = TRR$.

4) Tuning Important System Parameters: Figures 5-7 present the results on how the system's performance is impacted by several important system parameters, namely, the choice of sample distance computing algorithm, the number of best representative training samples used for classification (K), and the number of total training samples (M). Based upon these results, we tune the parameter values to balance the tradeoff between authentication accuracy and data collection/computing overhead.

Recall that our classifier employs a distance threshold of $(\mu + n\sigma)$, where μ and σ are calculated from the training samples. For each parameter study, we varied the value of n from 0.0 to 10.0 with an increment of 0.1, and had a total of 100 threshold values. We then plotted the TPR on y-axis and FAR on x-axis for each threshold value, resulting curves referred to as Receiver Operating Characteristics (ROC) curves.

Firstly, Figure 5 compares the performance of three distance computing algorithms: COS (cosine), COR (correlation) and DTW, assuming a single best representative training sample $K = 1$, music cue duration of 10s, and 30 training samples $M = 30$. Among these three algorithms, DTW fares much better than the other two: its EER is 14.29% smaller than that of COS distance, and 20.51% smaller than that of COR distance. This is as expected because DTW is designed to match the waveform of two signals [2] and thus outputs more accurate distance score. As a result, in the remaining of this study, we will use DTW for evaluation. Even though DTW incurs more computation than the other two, our Google Glass implementation shows that through software optimization, the processing latency of DTW distance can be made very small (see Section IV-C).

Secondly, Figure 6 compares the performance of two K values: K=1 and K=3, assuming DTW distance, music cue duration of 10s, and 30 training samples. Recall that our classifier compares the test sample against K best representative training samples, generates K independent classification results, and votes among them for the final authentication result. Hence, we expect that considering top 3 samples will be better than only considering the top 1 sample, as confirmed by the results shown in Figure 6. However, we observe the improvement is very marginal: the EER when $K = 3$ is only 0.5% smaller than the EER when $K = 1$. On the other hand, having $K = 3$ incurs three times as much computation as having $K = 1$. As a result, in the remaining of this study, we will use $K = 1$ for evaluation.

Thirdly, Figure 7 compares the performance of three training dataset sizes: M=10, 20, and 30 samples, assuming DTW distance, $K = 1$, and music cue duration of 10s. We observe that the EER when $M = 30$ is 1.79% smaller than the EER when $M = 20$ and 3.17% smaller than that when $M = 10$. We emphasize that the value of M has NO impact on the computing overhead in the online authentication phase because the classifier only compares the test sample against K representative training samples. As a result, in the remaining of this study, we will use $M = 30$ for the evaluation. Please note that we don't choose use a larger M value because the benefit of having a larger M diminishes quickly after having 30 samples.

5) Authentication Accuracy Results: After tuning system parameters to balance accuracy and computing overhead, we next calculate the EER value of the resulting *Headbanger* system when users choose different music cue durations and present the results in Figure 8. As soon as a user starts the authentication procedure, how long the user must wait before she receives the authentication result is an important quality-of-service metric, which we refer to as *authentication latency*. Authentication latency consists of two parts: data input latency and data processing latency, wherein the former is the time a user spends on listening to the music cue and making head movements while the latter is the time *Headbanger* spends on computing the authentication result. Between these two parts, data input latency is by far the bottleneck as the data processing latency can be easily reduced (by better algorithms and/or

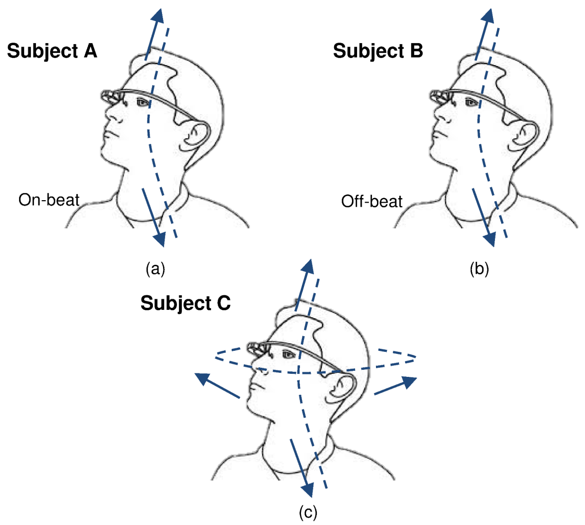


Fig. 9. Pictorial description of the nodding pattern employed by each target. Target (a) only moved his head in the vertical direction, and his nodding immediately follows each music beat (on-beat); target (b) only moved his head in the vertical direction, but there is often a delay between his nodding and the music beat (off-beat); target (c) occasionally combined shaking with nodding, and on-beat. As a result, the nodding patterns in (b) and (c) are slightly more complex than that in (a).

faster hardware) and/or hidden (by pipelining computing with data collection). Unfortunately, data input latency is hard to be reduced or hidden by the improvement in software or hardware. Recognizing that different users can tolerate different latencies and desire different levels of authentication accuracy, *Headbanger* allows the users to choose the music cue duration (which has the same length as the data input latency).

From Figure 8, we observe that the EER value of *Headbanger* is 6.65% with a 5-second music cue, and 4.43% with a 10-second music cue. We take the viewpoint that such error rate is rather sufficient for personal head-worn devices that are not used in hostile environments. Also, the data input latency of 5-10 seconds is comparable with similar authentication systems. For example, a gait-based authentication system [13] delivers an EER of 7.3% after observing the user for 20 meters; a pattern-swipe system [7] delivers at best a TPR of 98% and a FAR of 50% with offline processing; an eye-blinking-based authentication system [31] delivers the BAC=(TPR + FAR)/2 of 94.4%, with a processing latency of 32 seconds.

B. *Authentication Robustness of* Headbanger

After evaluating authentication accuracy for *Headbanger*, we next study its robustness. In this study, we focus on imitation attacks. For this purpose, we asked a number of participants (attackers) to imitate simple nodding patterns from three participants (targets) after watching the target's video for as long as they desire, and calculated their chances of successful imitation - i.e., the attacker is mistakenly accepted by *Headbanger*. We note that nodding is the simplest head-movement pattern and easiest to imitate; hence, the results presented here represent the *lower bound* robustness that *Headbanger* offers. In reality, users are more likely to employ more sophisticated head-movement patterns, which will be much harder to imitate.

1) Participants: We had a total of 37 volunteer participants, including 31 males and 6 females. The average age of the participants was 25.6 years with a standard deviation of 6.6 years. The youngest participant was 22 years old while the eldest was 49 years old.

2) Procedure: Our second experiment aimed at emulating a practical imitation attack scenario. In this experiment, three targets recorded video when they were nodding with a music cue. Note that the music is usually played via a bone conduction speaker or an earplug, and that it is difficult to use a camcorder to capture the music sound in a noisy environment. To address this concern, during recording, we set the speaker volume to maximum and conducted the recording in a quiet laboratory environment.

We divided the attackers into three groups, and asked each group to imitated one target. In each session (consisting of 30 trials), the attacker could watch the video for as long as they wish. Our system provided a feedback after each trial so that the attackers could adjust their nodding pattern if they wanted. After the attacker had 30 trials, we ended the session no matter whether the attacker had succeeded or not. In each session, we noted the total number of successes the attacker had as well as the number of trails before the first successful imitation.

3) Results: Each of the three targets performed simple nodding in this experiment; though simple, their nodding patterns have varying complexity. As shown in Figure 9, Target A moved his head vertically, each nodding on a music beat, with no noticeable horizontal movement; target B also only moved his head in the vertical direction, but there was always a delay between his nodding and the corresponding music beat; target C combined slight shaking along with nodding, and closely followed music beats. Among these three targets, A is the easiest, while the other two are slightly more complex.

Table I summarizes the results. Here, we use FAR to denote the successful imitation rate because a successful imitation is counted as a false acceptance in our system. The overall FAR of the experiment is 6.94%, while the individual FARs for the three targets are 15.83%, 2.77% and 2.72% accordingly. Since target A had the easiest nodding pattern, 7 out of 12 participants could succeed at least once during their 30 trials, while for targets B and C, the numbers are 3 out of 13 and 3 of 12, respectively. These results are very promising: when a user employs slightly more complex head-movement patterns, it becomes much harder for others to imitate (EER dropping from more than 15% to around 2.7%).

Target	No. of Attackers	No. of Successful Attackers	Average No. of Trials Before First Successful Login	FAR (%)
A	12	7	10.33	15.83
B	13	3	14.33	2.77
C	12	3	17.67	2.72
Overall	38	13	13.17	6.94

TABLE I. The attack results show that as head-movement patterns become more complex, it becomes much harder to imitate.

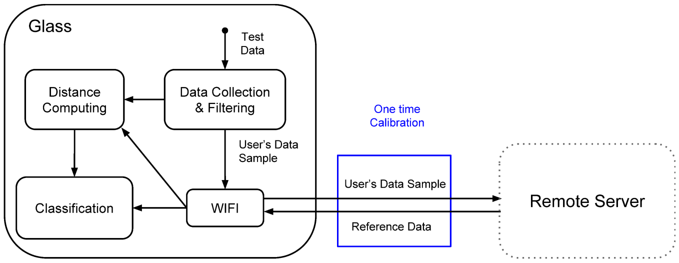


Fig. 10. Software modules of *Headbanger* implementation

C. *Headbanger Google Glass App Implementation*

In the second phase of evaluation, we implemented *Headbanger* on Google glass as an authentication app. Figure 10 shows the main software modules in the app. Upon initiation by the user, the app plays a music cue for a user-specified duration. The user conducts head-movements in synchrony with the music cue while the app records the accelerometer data in parallel. At the end of the music cue duration the app enters the data processing phase where the sensor readings are input to the *Headbanger*'s software modules for processing. Upon completion of data processing, the app responds with a YES or NO textual output on the Google Glass screen. The training phase is conducted offline on a PC, and we ensure that the training set is readily available during the authentication process.

1) Data Processing Latency: In Table II we report the measured average processing latency of the *Headbanger* app for music cue durations of 5, 6 and 10 seconds. The processing latency is within 2 seconds for a 10-second data input, and is less than 1 second (0.88s) for a 5-second input. We also report the total latency breakdown for different software modules, and find that DTW computation is the main bottleneck. In our implementation, we used Fast DTW [33] which decreases the computation complexity from $O(n^2)$ to $O(n)$ without compromising the authentication accuracy.

As we discussed earlier, the processing latency can be further reduced, and it could be partially hidden if we pipeline the data processing and data input. As a result, we believe that our design of *Headbanger* is indeed *light-weight* and it is realistic to run the *Headbanger* app on devices that have comparable computing capabilities as the Google Glass.

V. Related Work

Several studies have looked at head or eye movements for various purposes including user authentication. Harwin et al. [17] are the first to use head gestures for human computer interaction. Westeyn et al. [37] used eye-blinking pattern as a unique feature for authentication. They achieved 82.02% accuracy with 9 participants. Rogers et al. [31] proposed to use unconscious blinking and head movement to identify a user from a group of users. In this method, users were asked to view rapidly changing pictures for 34 seconds before they can be identified. Ishimaru et al. [21] comes close to our system design; they proposed to combine the eye blinking frequency from the infrared proximity sensor and head motion patterns from accelerometer sensor on Google Glass to recognize activities (e.g., reading, talking, watching TV, math problem solving). We note that recognizing activities is usually much easier than recognizing who is performing the activity, which is our objective in this study.

music cue duration (s)	data processing latency (s)	time breakdown (%)		
		Filtering	DTW	Thresholding
10	1.93	0.49	99.50	0.01
6	1.15	0.63	99.36	0.01
5	0.88	0.82	99.19	0.01

TABLE II. Measured response time of *Headbanger* app implementation on Google Glass with different music cue durations and for $K = 1$. The response time reported here is an average over 20 trials.

There are also a number of physiological activity recognition studies using computer vision methods [18], [26]. While [26] primarily uses computer vision to detect head gestures, Bio-Glass [18] combines Google Glass's accelerometers, gyroscope, and camera to extract physiological signals of the wearer such as pulse and respiratory rates. Camera processing on wearable devices, especially Google Glass is compute intensive and has a high energy budget [27].

Accelerometers have long been used to sense, detect and also recognize movements in other parts of the body; for example, gait recognition requires sensing in areas such as waist [1], pocket [14], arm [15], [29], leg [13] and ankle [12]. These techniques, though well known, may not be suitable for on wearable devices due to complexity (computation and energy) in the machine learning process.

Hand gesture and touchscreen dynamics are often coupled for authenticating to a (touchscreen) device (see e.g. [5] for a survey). A number of features [5], [32] (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 for authentication purposes [5], [9], [11], [34].

VI. Concluding Remarks and Future Direction

As wearable devices are increasingly weaved into our everyday life, providing security to the data acquired by or accessed through these devices becomes critically important. In this study, we have developed a user authentication system that uses head-movement patterns for direct authentication to a wearable device. Compared to existing authentication solutions, the proposed solution delivers accurate authentication, is robust against imitation attacks, incurs low processing delays, and offers great convenience to users.

Through an extensive evaluation that involves 95 users, we observe that the average true acceptance rate of our approach is at 95.57% and the false acceptance rates at 4.43%. We also observe that even simple head-movement patterns only allow less then 3% of the imitation attacks to succeed. We have also implemented an app on Google Glass, and measured the end-to-end processing delay of less than 2 seconds for a 10-second data sample. As a result, we believe it is realistic for the proposed authentication system to be executed on resource-constrained devices such as smart-glass. We further believe the proposed method can help enable wider deployment of wearable devices and apps in our life. Towards this goal, in our future work, we will focus on how to make the head-movement based user authentication approach more reliable in real-world settings.

VII. Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant Number 1228777. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] 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.

[2] D. J. Berndt and J. Clifford. Using dynamic time warping to find patterns in time series. In *ACM KDD AAAI workshop*, 1994.

[3] 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.

[4] R. Challis and R. Kitney. The design of digital filters for biomedical signal processing part 3: The design of butterworth and chebychev filters. *Journal of biomedical engineering*, 1983.

[5] G. D. Clark and J. Lindqvist. Engineering gesture-based authentication systems. *Pervasive Computing, IEEE*, 14(1):18–25, 2015.

[6] C. Cornelius, R. Peterson, J. Skinner, R. Halter, and D. Kotz. A wearable system that knows who wears it. In *ACM MobiSys*, 2014.

[7] 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.

[8] A. De Luca and J. Lindqvist. Is secure and usable smartphone authentication asking too much? *Computer*, 48(5):64–68, May 2015.

[9] 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.

[10] Fitbit. http://en.wikipedia.org/wiki/Fitbit.

[11] M. Frank, R. Biedert, E.-D. Ma, I. Martinovic, and D. Song. Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *Information Forensics and Security, IEEE Transactions on*, 8(1):136–148, 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*, 1(7):51–59, 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] Google glass. http://en.wikipedia.org/wiki/Google_Glass.

[17] W. Harwin and R. Jackson. Analysis of intentional head gestures to assist computer access by physically disabled people. *Journal of biomedical engineering*, 12(3):193–198, 1990.

[18] 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.

[19] R. Hoyle, R. Templeman, D. Anthony, D. Crandall, and A. Kapadia. Sensitive lifelogs: A privacy analysis of photos from wearable cameras. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1645–1648. ACM, 2015.

[20] R. Hoyle, R. Templeman, S. Armes, D. Anthony, D. Crandall, and A. Kapadia. Privacy behaviors of lifeloggers using wearable cameras. In *ACM UbiComp*, 2014.

[21] 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.

[22] A. K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(1):4–20, 2004.

[23] S. Jana, A. Narayanan, and V. Shmatikov. A scanner darkly: Protecting user privacy from perceptual applications. In *Security and Privacy (SP), 2013 IEEE Symposium on*, pages 349–363. IEEE, 2013.

[24] A. R. Jensenius. Action-sound: Developing methods and tools to study music-related body movement. 2007.

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

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

[27] R. LiKamWa, Z. Wang, A. Carroll, F. X. Lin, and L. Zhong. Draining our glass: An energy and heat characterization of google glass. In *Proceedings of 5th Asia-Pacific Workshop on Systems*, page 10. ACM, 2014.

[28] F. Monrose and A. D. Rubin. Keystroke dynamics as a biometric for authentication. *Future Generation computer systems*, 16(4):351–359, 2000.

[29] 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.

[30] T. Rahman, A. T. Adams, M. Zhang, E. Cherry, B. Zhou, H. Peng, and T. Choudhury. Bodybeat: a mobile system for sensing non-speech body sounds. In *ACM MobiSys*, 2014.

[31] C. E. Rogers, A. W. Witt, A. D. Solomon, and K. K. Venkatasubramanian. An approach for user identification for head-mounted displays. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pages 143–146. ACM, 2015.

[32] 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.

[33] S. Salvador and P. Chan. Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 2007.

[34] 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.

[35] Smart watch. http://en.wikipedia.org/wiki/Smartwatch.

[36] S. V. Stevenage, M. S. Nixon, and K. Vince. Visual analysis of gait as a cue to identity. *Applied cognitive psychology*, 13(6):513–526, 1999.

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

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