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

ABSTRACT

In this paper, we present the design, implementation and user studies of a novel approach to authenticate wearable devices users based on their unique behavioral patterns. We prototype an authentication system, dubbed *Headbanger* for head-worn wearable devices by monitoring user's unique head-movement patterns in response to an external audio stimulus. Solutions today primarily rely on indirect authentication mechanisms through the user's smartphone, which can be cumbersome and more susceptible to adversary intrusions. Biometric solutions, are subject to the availability of the specific sensors in the wearable unit. Using a head-worn personal imaging device as a running example and through extensive experimental evaluation with 30 human subjects, we show that our mechanism can authenticate users with an average acceptance rate of 95.1% while keeping the average false acceptance rate of 3.9%.

INTRODUCTION

Wearable devices are now available off-the-shelf and on the way to become an integral part of human lives [?, ?, ?]. This is thanks to the advances in hardware miniaturization technology, affordable sensors and processor chips, and low-power computing. The wearable devices typically collect data about their wearer and their surroundings. This collected data on such devices is personal in nature and often relates to the user's health. There has been work on limiting privacy threat to other users [?, ?, ?]. Any security solution for these devices has to strike an appropriate balance with user convenience, especially as users are interacting with an increasing number of such specialized devices. A fundamental building block for safeguarding the security of user data acquired on or accessed through wearable devices are user authentication techniques.

Authentication Challenge. Authentication on most commercially available wearable devices today [?, ?] relies on an indirect mechanism, where users can log in 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 devices. The security of this approach is also in question as it increases the chance of hacking into both the devices if either of those are lost or stolen. Some devices including Google Glass [?] and FitBit's health tracker [?] allow linking the device to online accounts instead of the phone for user's convenience. This, however, does not add any security. Indirect authentication remains a dominant paradigm for wearables despite these fundamental shortcomings because these devices are *seriously resource-constrained* in many aspects: battery power, computational and storage capabilities, and input/output methods. As a result, typical authentication methods designed for more powerful devices can not be directly applied and must operate indirectly through a paired smartphone or other more capable device. In this paper, however, we 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 lightweight.

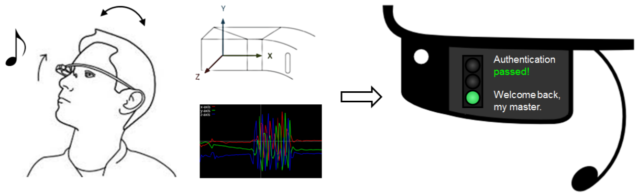


Figure 1. Illustration of Headbanger. The head-worn device authenticates the users based on signatures generated from head-movement patterns. These patterns are created in response to an audio snapshot played on the device.

Before we explore direct authentication methods for wearable devices, let us first consider the 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 quite 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, etc.) requires specialized sensing hardware and processing resources that add cost, and many of these sensors are 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 *behavioral* biometrics, and existing work has largely studied it in the context of authenticating smart phones and tablets [?, ?, ?, ?, ?, ?, ?, ?].

The main advantage of using behavioral biometrics for mobile devices is that the signatures can be readily generated from raw data of built-in sensors such as motion sensors, camera, microphones etc. Considering that cameras and microphones, as well as vision and audio processing algorithms, are quite energy-hungry, we thus focus on those behavioral biometrics 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 one type of behavioral characteristics: 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 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 light-weight authentication system. A key distinguishing feature of our work is that we will also consider stimuli that wearable devices can provide to design challenge-response inspired mechanisms, 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 generates a signature from user's head-movements. These signatures are used as the behavioral biometric. 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. Although 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, artificial intelligence devices.

In summary, the key contributions of this paper are:

1. We have designed and implemented a novel user authentication method to 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 behavioral biometrics for authentication. We design a 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.
2. Through comprehensive experiments involving multiple users and over different system design parameters we show that head-movement patterns can be used as a behavioral biometric. Our approach effectively identifies a wearable device user, with average false acceptance rate of 3.9% and an average true-positive rate of 95.1%.
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 response time of 4.4 seconds on the Google Glass for the most accurate results.

BACKGROUND

Wearable Device Authentication

Biometrics allow a system to identify a user based upon "who you are" (i.e., her physiology) instead of "what you have" (i.e., ID cards) or "what you know" (i.e., passwords) [**?**, **?**, **?**]. Physiological biometrics such as DNA, ear shape, face, fingerprint, hand or 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 [**?**]. Even characteristics such as body weight and fat percentage have been considered as secondary biometrics for authentication purposes [**?**].

However, biometrics are not prominently used in wearable devices that are commercially available today, though there have been specific point commercial designs (e.g., Nymi [**?**]). This can be attributed to the fact that biometrics would require the specific hardware and sensors 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 [**?**]. Okumura et.al. [**?**] have shown that the human arm swing patterns can be used to create signatures to authenticate to their cell-phones. Monrose et.al. [**?**] 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 [**?**] and touchpad touching dynamics [**?**, **?**] 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 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.

Head-movement as a Biometric

According to Jain et al. [**?**], 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 collecting head-movement patterns using the built-in accelerometers and motion sensors has become more accessible. Such sensors are available on almost all head-worn wearable devices available today, thus making head movements, both, available *universal* and *collectible*.

In this paper, we will show that free-style head movements are *distinctive* and *repeatable*, especially when combined with external stimuli such as music. 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). It has been shown [**?**] 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. Most adults naturally perform head movements or hand movements when listening to a fast beat audio track. 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.

Based on a preliminary analysis of the accelerometer signals from five Google Glass users, we also observed (see Figure **??** (a)-(e)) that these users *repeatedly* showed unique and *distinctive* head-movement patterns (that are differentiable through simple signal processing techniques), 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.

FEATURE

In order to explore the insightful information of musical head movement, we extract several feature from the original accelerometer data:

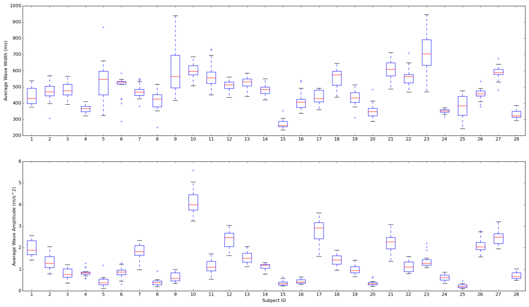


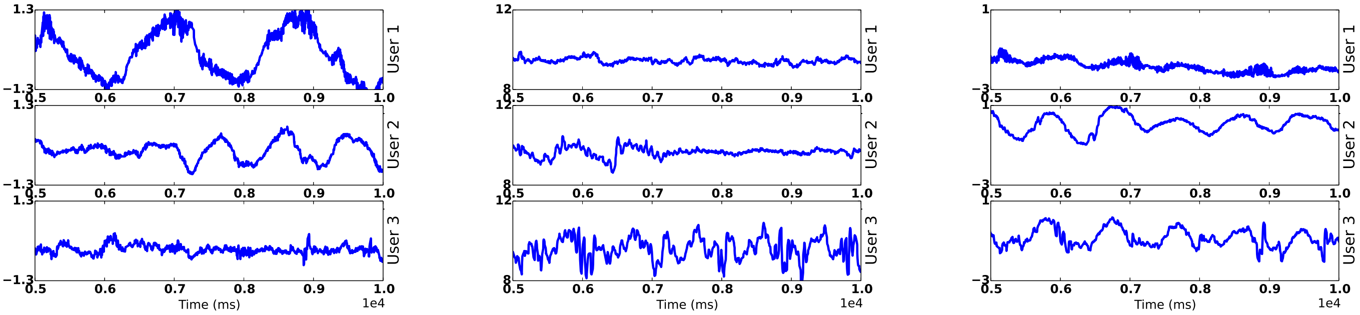
Figure 3. Wave Amplitude and Width Boxplot

- *Wave amplitude*,which is the difference of a wave bottom and its next wave peak. It measures the acceleration when the user performs the head movement, in other words, wave amplitude describes how hard the user nod his head. Due to accelerometer data usually contains high frequency noise, and we are more interested in the main signal where user nod his head, a 5-Hz-cutofflow-pass filter is applied before we detect peaks and bottoms in the signal .

- *Wave widh*, which is the time interval between two bottoms or two peaks. It describes the time to complete one nodding movement. The user is stimulated by a music cue, wave width is affected by the time interval between beats.

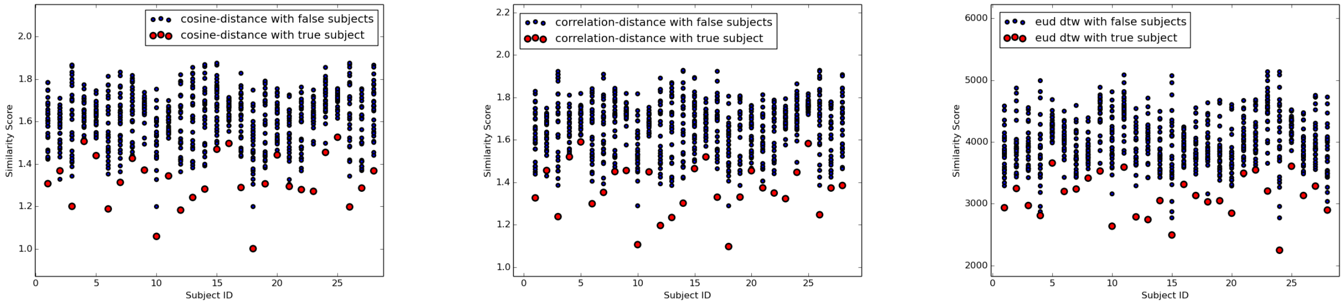
- *Series of response time (SRT)*,which is the series of time interval between the music beat and corresponding movement. Studies in [**?**] and [**?**] show this feature could be used to differentiate the users. Note that user is not necessary responding after the beat, as he can anticipate if he is familiar with the music cue. Thus, response time in our measurement can be either negative or positive. However, finding the associated movement is a non-trivial process, hence we develop the following algorithm for a robust detection. First, accelerometer data needs to be synchronized with the music cue in midi format. Note that midi file is a time series of instrument commands, no additional signal processing is required to detect the beat in the music as we can take each command as a beat if the music is simple enough. Second, we locate two neighboring peaks where a beat resides in between, and compute the time intervals from the beat to both peaks. Finally, we choose the one with a smaller absolute value as the response time. In some cases, user is not necessary nodding for every beat, thus we use the 90 percentile of the response time sample to interpolate when the detecting response time is beyond this value.

As shown in Figure **??**, most subject s wave amplitude are with different means and small variance compared with their wave width. To better understand the SRT for different users, we apply various similarity scores to provide an intuitive and quantitative description of their difference. In Figure **??**, Cosine Distance (COS) and Correlation (COR) can help most of the subject to differentiate himself with the other subjects, but for some subjects's true subject score (red dot) are closed



(a) X-Axis (b) Y-Axis (c) Z-Axis

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 the same music track wearing the same Google glass. The plots indicate that different users' head movement patterns appear distinctive from each other. The three users wore a Google Glass (in turns) and listened to a 10 second audio snapshot of a pop song.



(a) Cosine Distance (b) Correlation (c) DTW

Figure 4. The similarity score is computed over N (N = 20) sample data for each subject, and 28 subjects in total. For each column, a red dot represents the average similarity score of that subject's own SRT, i.e., each of his sample data compared with the other N-1 SRTs from him, hence we have the average similarity score over all these comparison. Similarly, a blue dot on the same column represents the similarity score of that subject's SRTs compares with another subject's SRTs.

to their false subject score (blue dot). For subject 2, red dot even exceeds the blue dot.The output Dynamic Time Warping (DTW) [**?**] can be used as another distance metric. ***describe why DTW is better than the other two in this case***. In Figure **??**s, it shows that all average self-comparing similarity scores are lower than that when they are compared with others. The observation above indicates the possibility to further derive a system which can differentiate the true user and the false users based on the fusion of these features.

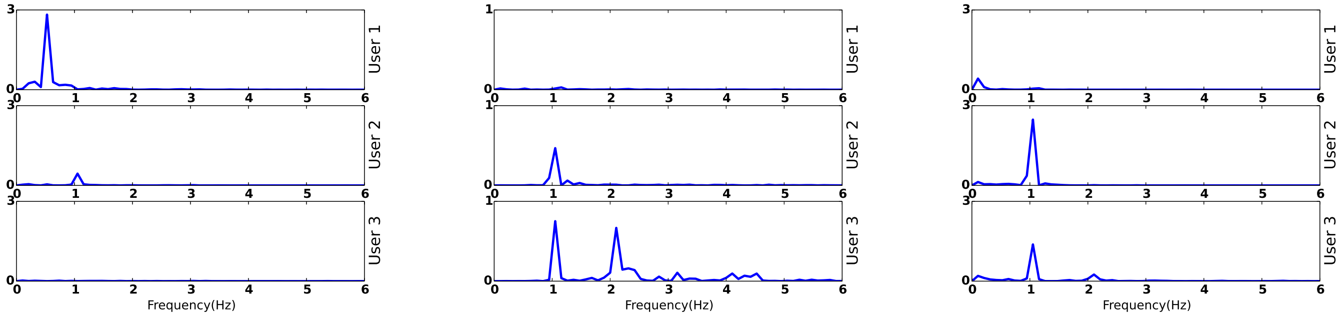
HEADBANGER SYSTEM DESIGN

In this work, we design a system, *Headbanger*, that will enable direct authentication of users to their smart-glass device using head-movements. The authentication process has two phases: a offline training phase and an online authentication phase. In the training phase, the system collects the real user's head movement data and use them as conducts the training to extract the features. We refer to this set of features extracted through the training process as the *trained* dataset which essentially serves the purpose of a reference in the matching stage. 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 will be possible with minor modifications, namely, by appropriately indexing the users in the trained database.

In the online authentication phase, the sensor readings during the authentication attempt are processed, features are extracted, matched with the trained set. The user is authenticated upon a successful match. We posit that *Headbanger* 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 [**?**]. The authentication process is initiated by playing a short duration audio track on the device, to which the user responds through head-movements. 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. We will refer to the audio snapshots as *music cues* in the rest of the paper. *Headbanger* generates unique features from the head movements of a user, and uses them as a unique signature for identifying the right user of 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 **??**, the user authentication in *Headbanger* involves the following key processes:

- *Sensor data collection*: *Headbanger* records the head-movements in the form of raw accelerometer signals (in 3



(a) X-Axis (b) Y-Axis (c) Z-Axis

Figure 6. Accelerometer data from three users, in the frequency domain. The data show that the spectrum is significantly concentrated within 5Hz for all three users.

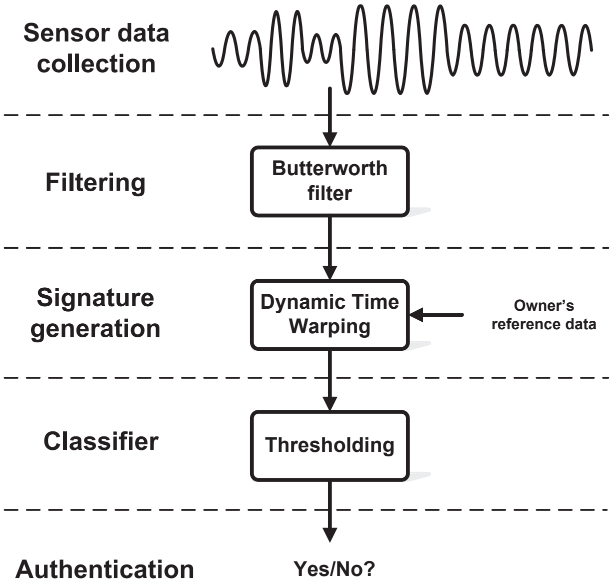


Figure 5. *Headbanger* system design flow. The online authentication phase of *Headbanger* consists of the following steps: (1) sensor data collection in which we collect accelerometer data while users move their head as a response to an audio track played on the glass, (2) filtering in which we apply a Butterworth filtering to smoothen the sensor data for subsequent processing, (3) signature generation in which we calculate the dynamic time warping (DTW) distances between two accelerometer samples as the signature, and (4) classification in which we adopt an adaptive thresholding mechanism to classify the user's head movement, whose result will be used as the authentication result.

dimensions) using the inbuilt accelerometer sensor on the smart-glass device.

- *Filtering*: The accelerometer signals are filtered by applying a 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.
- *Classification*: The signatures are classified as a match or not a match, based on a thresholding scheme and using the

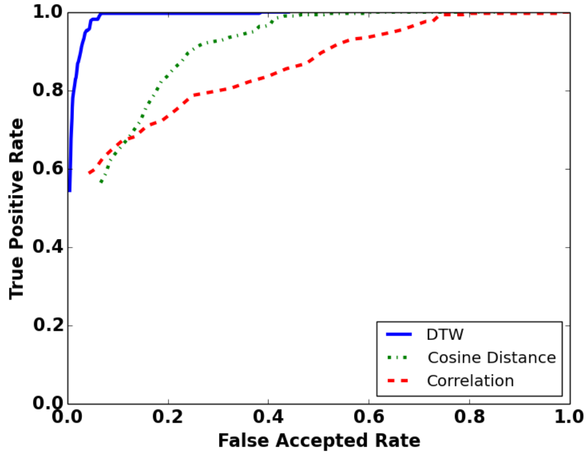


Figure 7. Evaluation of impact of different distance metric (DTW, cosine distance, and Correlation). Although DTW is relatively computing-intensive, ROC curve indicates that DTW provides a large enhancement over the other two metrics.

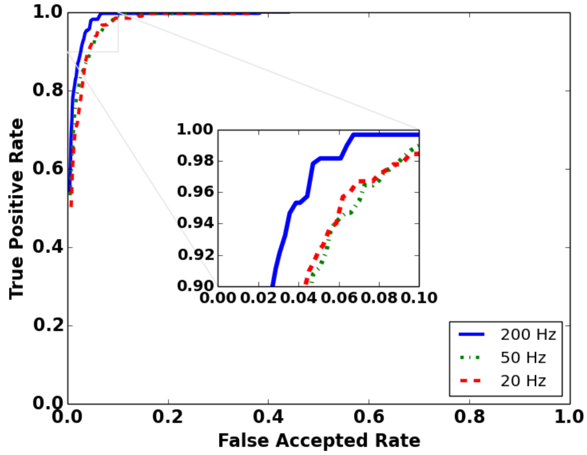


Figure 8. Evaluation of impact of different sampling rate shows that the highest sampling rate 200 Hz gives the best resultt. However, the sampling rate determines computational effort for the smart device, which could be significant in terms of response time.

trained data set as a reference. The system grants the user access to the device if there is a match with sufficient confidence.

We will now discuss these design aspects in more detail.

Sensor Data Collection

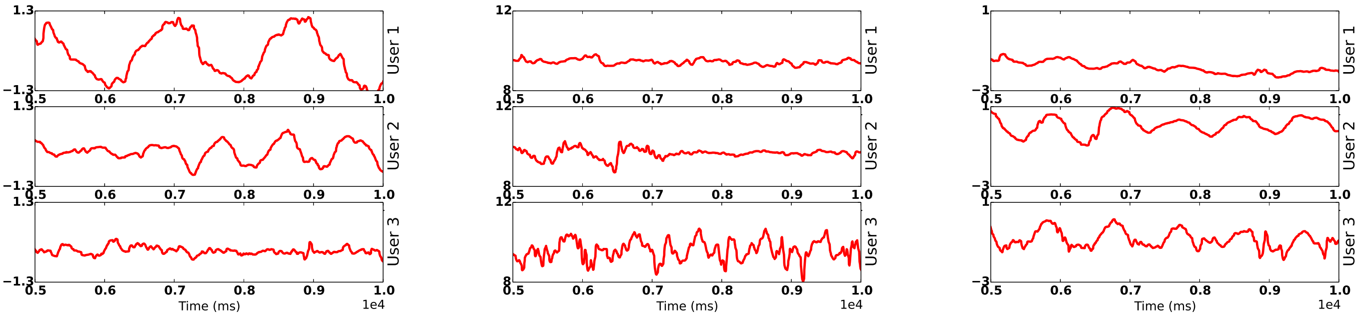
The sensor data collection step involves the user wearing the head-worn device and making head-movements in response to the music beats played on the device for a stipulated duration of T seconds. In this duration, the raw accelerometer signals, from the inbuilt sensor, are collected at a sampling rate of r samples/sec; the default sampling rate on Google Glass is 200Hz. The accelerometer data corresponding to one user, is a collection of accelerometer readings on the 3D axis (x, y, and z) collected over T sec duration. Figure **??** illustrates the axis conventions with respect to the user's head position. The data collection unit stores the accelerometer readings in a matrix with dimensionality $3 \times rT$, where each element corresponds to one signal point. We will refer to this $3 \times rT$ as a *sample* in our design. 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. This intuition will be more clear from the filtering stage to be discussed next.

Filtering

The accelerometer 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 frequency spectrum of each accelerometer sample, as shown in Figure **??** for three 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 that we used in our experimentation was contained 94 beats per minute [**?**]. 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 accelerometer 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 **??** (a)-(c) show the filtered results of the raw accelerometer samples shown in Figure **??**; compared to raw data, the filtered results are much more suitable for subsequent processing.

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 (temporal) signal over a certain time-window and yields a distance measure as the score. A low score (DTW distance) implies that the test signal is in close match with the reference. We use the DTW to generate a signature for the head-movements from the accelerometer



(a) X-Axis

(b) Y-Axis

(c) Z-Axis

Figure 9. Filtered accelerometer signals. Applied Butterworth filter of order 2 and cut-off frequency 10Hz.

signal. We observed from our preliminary tests that, users often start head movement at an angle with the vertical which varies among users. However, we also observed that the head-movements that follow exhibit a consistent, and often periodic, patterns over time. Treating the accelerometer sample set from the first trial as a reference, we apply the DTW algorithm on the successive accelerometer 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 accelerometer pair we generate the mean-value DTW distance, which is treated as the head-movement signature or unique *feature* for that particular user and audio combination.

In the offline training phase, we conduct M trials of head-movement exercises, collect M training samples, and obtain $M-1$ reference distance vectors. We observed from our evaluations (to be discussed in the next section) that $M = 30$ can yield in high accuracy while $M = 10$ can yield reasonable accuracy. The trade off is the computation overhead that goes into conducting the training (primarily DTW computations) for the M samples.

Classification

The classification step labels a test sample based upon the pre-established training/reference samples.

In this study, 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 M training samples, we identify the top K samples that have the lowest K average distance to the rest of the training set, and calculate their DTW vectors to the rest of the samples in the training set to obtain $(M-1)K$ distance vectors. We call this resultant vector as Top-K reference distance vector. For example, if $K = 1$, then we call the resulting algorithm as Top-1 algorithm; if $K \neq 1$, then we call the resulting algorithm as Top-K voting algorithm. The voting refers to the procedure that the DTW computation results for each sample in the training set is referenced, sorted (in increasing order of DTW scores) and the classification (a binary index, match or no-match) is performed on the top K entries. The final classification result corresponds to the majority vote from the list of K results. By using top K samples, instead of all M samples in the training set, we significantly reduce the computation overhead in the authentication process. In our evaluation, we will study the impact of K value; in particular, we evaluate for $K = 1$ and $K = 3$.
2. Compute the 3 element (x,y,z axis) mean μ vector and standard deviation σ vector of the top-K distance vector, for each authentication trial. We define "true" range as $[\mu - n\sigma, \mu + n\sigma]$, where n 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 $[\mu - n\sigma, \mu + n\sigma]$ as the threshold in our classifier design, and the strictness of this threshold is characterized by the value of n; a large n relaxes the threshold but can increase the false acceptance rate while a strict threshold with small n value can result in a high rejection rate of true samples. In our system we aim to reach an optimal value of n that can result in acceptable accuracy.

If the user's data is classified as "true" then the user is authenticated to the device; otherwise, the user is rejected.

EVALUATION

We evaluated *Headbanger* with comprehensive laboratory studies involving human subjects. We collected from volunteer participants accelerometer sensor readings with Google Glass. We analyzed these traces offline on a PC. Our evaluations are primarily aimed at determining the accuracy of detecting and differentiating users based on their head-movements, and understand the effect of design parameters such as length of the music cue and training data-set size, on accuracy. We also measured the response time of our Google Glass implementation of *Headbanger*. Our studies were approved by the Institutional Review Board (IRB) of our institution.

Method

Participants

We had total of 30 volunteer participants. The participants list included a total of 19 males and 11 females. The average 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 at 54 years.

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 spectacles were asked to remove their spectacles 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. We conducted these trials for three duration values: 5 s, 6 s and 10 s. As we will show further, the accuracy of the system can significantly improve with the duration of the music cue; longer the duration better is the accuracy. 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.

During the course of the experiment, the participants were allowed to take a break or withdraw from experimentation if they felt uncomfortable at any point; for example, feeling dizzy after head-movement for a period of time, not being able to see clearly if near-sighted, etcetera. The conductor also allowed the user to take a break of about one minute after each experiment trial. Each trial lasted for the duration of the music cue played on the Glass, and a total of 40 such trials were conducted for each of the 30 subjects. The experiment lasted over a duration of 60 days, of which 15 subjects 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 an average per subject. The experiment yielded three sets of data traces that each correspond to the three music cue durations we selected.

Our second experiment aimed at an practical imitation attack field test scenario. In this experiment, there were three subjects being taken video while they were performing their login movement. The participants were divided into three group, and each group was asked to imitated one of the three users above. During the test, the participants can watch the video for as many time as they want at anytime between two trials. A feedback from our system will be provided after each trail so that the participants can decide whether they need to adjust their movement. After total number of trials reached 30 for each participant, the experiment will stop no mater whether the participant succeed or not. This experiment is conducted in quite space on campus. We will now discuss our evaluation results for both experiments in detail.

Accuracy

We evaluate the accuracy of *Headbanger* using metrics that are commonly used in evaluating authentication systems, namely, the false acceptance rate FAR (percentage of false test samples that are mistakenly accepted), false rejection rate FRR (percentage of true test samples that are mistakenly rejected), and true positive rate TPR (percentage of true test samples that are correctly accepted). A strict threshold in the classifier can lead to high FRR, while overly relaxing the same can lead to a high FAR. Hence, we also consider the equal error rate EER (percentage of errors when $FAR = FRR$), that considers both FAR and FRR. Figures **??**, **??** and **??** report the accuracy of *Headbanger* through the metrics stated above. In general, our evaluation of the 30 subject data-set indicates that a TPR of 95.1% at FAR of 3.5%, and EER = 3.97% can be achieved in *Headbanger*, however, these results are tailored to the following parameter and algorithm choices: 10 sec music duration, $K = 3$, and 30 (out of 40) trials from each user being used for training with a thresholding parameter value $n = 2.7$ with DTW. We will now discuss the results and the impact of such parameter choices on accuracy in more detail.

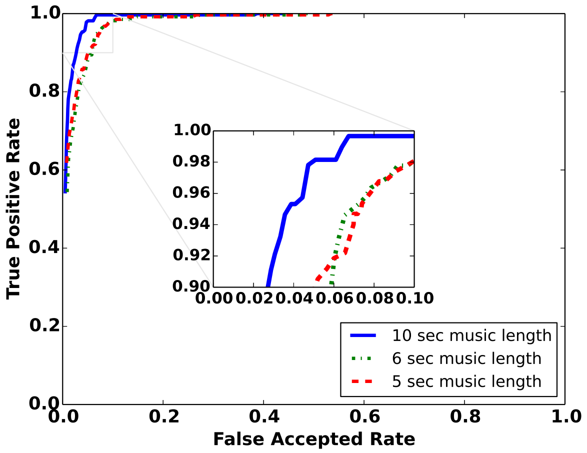


Figure 10. Evaluation of impact of music cue duration on TPR and FAR in Top-1 scheme ($K = 1$). A 10 sec music snapshot is trimmed into music cues of 10 sec, 6 sec and 5 sec correspondingly.The variable here is n. Each (TPR, FAR) data point in the curve corresponds to a different value of n

Impact of similarity algorithm

In previous preliminary study, we find that DTW, Cosine Distance, and Correlation are giving promising results, hence we will evaluate these three algorithms in our end-to-end system. In this experiment, we vary thresholding parameter value n and fix the other parameters: 10 sec duration, $K = 1$, and training data $size = 30$. We can observe from the ROC (receiver operating characteristic) curves in Figure **??**

Impact of music cue duration and K value

The classification algorithm in *Headbanger* generates the classification result (YES or NO) by voting among the individual results each generated by the top-K samples in the truing set. We can observe from the ROC curves in Figure **??** and **??** that for both, $K = 1$ and $K = 3$, the TPR is close to 95% while the FAR is slightly above 3% for the 10 sec music duration. For $K = 1$ the FAR increases to about 7% for the 5 sec and 6 sec cases, however, for $K = 3$ the FAR decreases to about 3% and 4%, respectively. We observe a similar trend with the EER as shown in Figure**??**, where improvements of 0.5 - 1 % can be achieved by choosing $K = 3$ over $K = 1$.

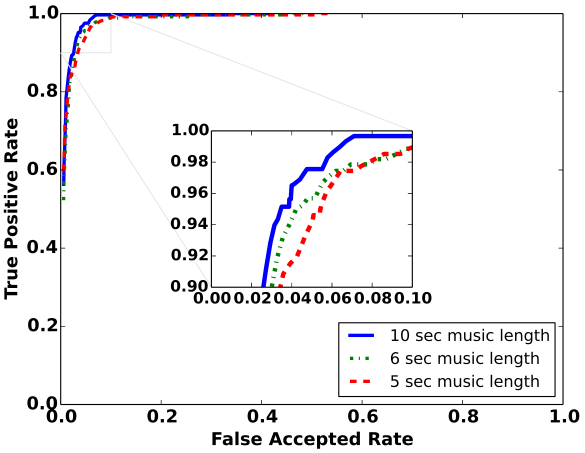


Figure 11. Evaluation of impact of music cue duration on TPR and FAR in Top-3 voting scheme ($K = 3$). A 10 sec music snapshot is trimmed into music cues of 10 sec, 6 sec and 5 sec correspondingly.The variable here is n. Each (TPR, FAR) data point in the curve corresponds to a different value of n

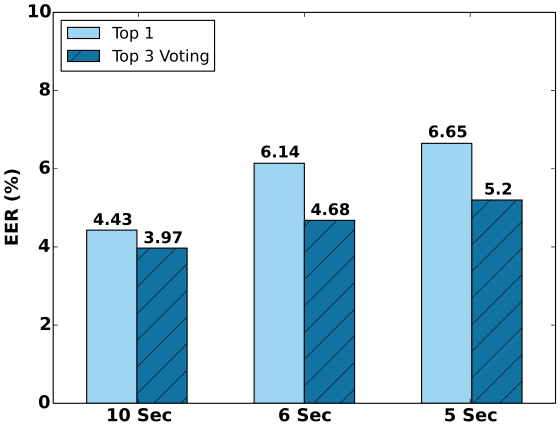


Figure 12. Comparison of EER for different music lengths (10 sec, 6 sec and 5 sec) with a fixed n value of 2.7

This indicates that the accuracy in *Headbanger* can improve with a larger value of K. However, the improvement in accuracy through redundancy in the training set trades off with the increased execution time as the matching requires at least $K - 1$ extra DTW computations as opposed to only 1 for a top-1 scheme. As we will show ahead, DTW computations incur heavy CPU budget on the wearable device.

In general, we observe from the results that the FAR can be decreased by increasing the music cue duration. We can observe (from Figures **??**, **??** and **??**) that the improvement is less significant when the music cue duration is increased from 5 sec to 6 sec, however, the improvement is more significant when the music cue duration is increased to 10 sec. In *Headbanger* the data collection phase for the authentication system (sampling accelerometer sensor readings) is executed in parallel with the music cue. The data processing phase involving the filtering, classification and matching is executed only at the end of the music cue and in the same order. We note that, the data input duration of 5-10 sec for authentication may seem long, but in reality, such data input durations are on par with those of password based systems [**?**].

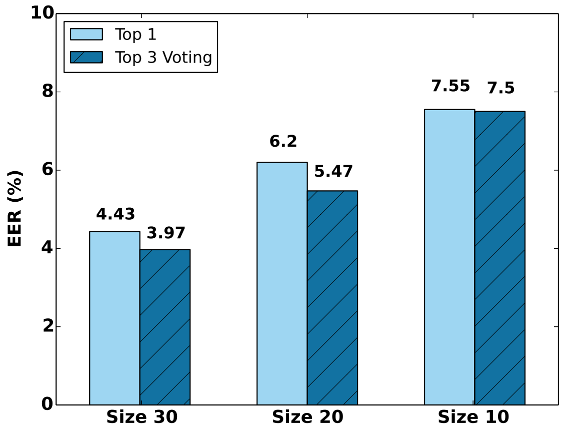


Figure 13. Comparison of EER for different training sample sizes (30, 20 and 10) with fixed n value of 2.7

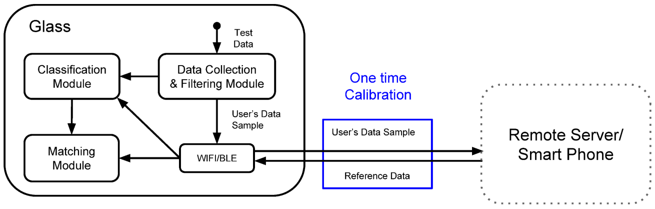


Figure 14. Software modules of *Headbanger* implementation

Impact of Training set size

Recalling from *Headbanger*section **??**, the input to the training phase is a set of temporal signals (samples with duration equal to the music cue duration), each corresponding to one trial of the head-movements from the user. Our evaluations so far considered a training set size of 30 samples. In Figure **??**, we report the EER in *Headbanger* for three different training data set sizes; 10,20 and 30 samples. We can observe from Figure **??** that the EER holds an inverse relationship with the training set size. A larger training set minimizes the variance in mean and standard deviation computations, as the errors in their inconsistency are reduced by averaging the mean and standard deviation estimates over a larger set of data. On the other hand, a larger training set also implies a longer execution time of the training phase. However, in our system design, we posit that the training phase can be conducted offline on a more compute efficient device (smartphone, PC or server) and that the wearable device can pre-fetch the trained data (for example, an XML file), prior-to or during data collection phase, through a wireless link.

Headbanger Google Glass App Implementation

music cue	response	time breakdown (%)		
duration (s)	time (s)	Filtering	DTW	Thresholding
10	4.4	0.20	99.80	<0.01
6	2.73	0.26	99.74	<0.01
5	2.44	0.28	99.72	<0.01

Table 1. 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.

Component	Power Consumption (mW)	Duration (s)
Sensor	29	10
Speaker	410	10
CPU	1600	14.4

Table 2. Power consumption on Google Glass of components relevant to *Headbanger*. The CPU (running at maximum frequency) power consumption includes that of the heads-up display screen being ON as well. Duration marks the time for which component was ON during the a 10 sec music cue length trial

We implemented *Headbanger* on Google glass, positioning it as an authentication application (app). Figure **??** shows the main software modules in the app. Upon initiation by the user, the app plays a music cue for a stipulated duration. The user conducts head-movements in synchrony with the music cue while the app records the accelerometer sensor in parallel. At the end of the music cue duration the app executes the data processing phase where the sensor readings are input to the *Headbanger*'s software modules for processing. The processing stage includes the filtering of the accelerometer sensor values, classification and feature extraction using DTW, and threshold based matching of the generated features with those from training set. Upon completion of data processing, the app responds with a YES or NO textual output on the Google Glass screen, depending on match score.

In our current implementation, the training phase is conducted offline, prior to live-testing off the application. The training phase involves collecting 30 samples (variable) of the head-movement accelerometer readings, generating the features, and saving them into a local server (running on PC) as an XML file, with appropriate indexing. Upon app initiation on Glass, the trained features are pre-fetched from the server through a wireless connection. This ensures that the training set is readily available during the authentication process, thus eliminating the additional processing time required for the training phase. Conducting online training, particularly that involves DTW computations, is very compute intensive on a resource constrained devices such as Glass. One possible solution would be for the Glass to offload the training phase computation to a local server machine.

Response time

In Table **??** we report the measured average response-time of the *Headbanger*implementation on Google Glass app for music cue durations of 5,6 and 10 seconds. We conducted the benchmark execution-time profiling of *Headbanger* on Glass in a controlled indoor laboratory setting with no mobility. We define response time as the time elapsed between music cue completion to the display of authentication response (YES/NO text) on the Glass screen. Our measurements indicate that the response time is within 5 seconds for a 10 second data input, and is almost halved for a 5 second data input. We feel that a response time of 2-5 sec for a local authentication solution in Google Glass is comparable to that of prior-art that comes close to our solution [**?**, **?**]. It is also important to note that authentication solutions that execute locally on head-worn wearable devices, especially on a heavily resource constrained device like Glass, are still not mature. However, the hope is that such solutions will possibly catch up to speed in the near future and that our approach is advancing one step in that direction. In Table **??** we also report the execution time of the key processes in *Headbanger*; filtering, DTW computation ($K = 1$ requires 1 DTW computation), thresholding based similarity matching. We can observe from Table **??** that the DTW computation dramatically compute intensive than the other processes.

It is important to note that our current implementation uses a faster version of the DTW algorithm called Fast DTW [**?**], providing about 2x speed-up in DTW computation. We believe that the response time can be reduced further through strategic methods such as, further optimizations in the Fast DTW algorithm or pipelining the app execution along with data collection. A specific strategy for reducing response time for rejected attempts can be that, after a short duration, before the entire music cue is played, if it is found that a user's movement does not match the signature of the claimed user with a sufficient pre-determined confidence level, then the onsite 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 online training and classification, to the user's Bluetooth paired smartphone or a nearby cloudlet [**?**].

DISCUSSION

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

Power consumption

Google Glass is an example of a wearable device that is heavily battery power constrained. Measuring the power consumption of the Glass's battery is a challenging task as that requires physically dismantling the device. We refer to the measurement paper on Google Glass by Robert et. al [**?**] for the power consumption of the key components relevant to *Headbanger* implementation on Glass: the speaker for music cue playback, the accelerometer sensor and the CPU being ON during the entire authentication process. We report the relevant numbers in Table **??**. While the high CPU power consumption may not necessarily be surprising, the speakers also extrude considerable energy from the battery. We note that one possible solution for future consideration would be to play the music cue as intermittent notes over the duration, for example a ping or a beat sound periodically, where the speaker would be switched ON only during playback.

Is this secure?

An authentication system must have an effective protocol ensuring security of the authenticating user's data. Our system runs an implicit authentication protocol where the user is given a finite set of (calibrated) music tracks to pick, based on which the user makes head-movements that are used as unique signatures for authentication. Our design assumes implicit security of the user's data, as a user voluntarily accepts the enforcement of conducting head-movements in response to the music. It is arguable that such enforcements are an integral part of most commonly used authentication systems; for example, typing a password, swiping the finger on the fingerprint sensor, approving of the camera recognizing the face. In all these cases the user is aware that he/she is inputing data into the system for authentication.

One way of compromising security would be a successful spoof of the head-movement by an adversary. For example, head-movements from an authorized user may be imitated by an adversary attempting to login to the device. If the head-movements from the user is regular (such as a nod), it may be easily imitated as opposed to a random head-shake such as a head-bang. To understand the effect of imitation on the accuracy of authenticating a user to *Headbanger*, we conducted an experiment (under the same set up as described in section **??**) where 29 volunteer participants were asked to imitate the head-nod movements of one user (one of the authors) who was trying to authenticate to the device using a 10 sec music cue. A total of 30 trials were conducted of which 10 samples were used as test data and 20 for training. Our evaluations of this dataset resulted in reasonable accuracy values of, an EER of 7.2% and a balanced accuracy $BAC = 1-((FAR+FRR)/2)$ of 94.5% for the authorized user. Our results indicate that attacking the system through imitation of a simple head-gesture can still be challenging.

Multi-Modality

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. The array of motion sensors (accelerometer, gyroscope, inertial measurement unit) open up opportunities for multi-modal motion sensing. For example, in Glass, 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 (IR) light sensor (we had to root the Google Glass to access the IR sensor unit) 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 [?, ?]. We reserve such potential enhancements to our system for future implementation.

Large Scale User Study

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 is typically infeasible in academic laboratory settings. We reserve such large scale experiments for future work, and hope to accomplish through industry collaborations. We will, however, be releasing our data-sets to the public in the near future.

RELATED WORK

Headbanger is a behavioral biometric based authentication system, which focuses on head movement. Below, we review the related literature on mobile device authentication.

The work by Harwin et al. [?] is usually considered the first to propose use of head gestures by combining pointing and movements for human computer interaction. In [?], the authors used eye-blinking pattern as a unique feature for authentication. They achieved 82.02% accuracy with 9 participants. Compared to eye blinking pattern, head-movements can provide much more entropy, therefore can be considered as a more suitable biometric characteristic. The work by Ishimaru et al. [?] 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) The key difference of our approach from [?] is that, their approach focused on common head-movement and eye-blink patterns when people employ the same activities such as reading, typing, etc. We carefully investigated the head-movements from human subjects and found that they are unique to each person. Our system also identified these head-movements with a higher accuracy (95% versus 82% in [?]).

There are also a number of physiological activity recognition studies using computer vision [?, ?]. While [?] primarily uses computer vision to detect head gestures, BioGlass [?] 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 [?].

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 [?], pocket [?], arm [?, ?], leg [?] and ankle [?]. 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. A number of contextual features including biometrics [?] (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 [?, ?, ?]. While most of the techniques require users to explicitly conduct a gesture following a specific pattern, TIPS [?] proposed a multi-stage filtering with dynamic template adaptation strategy to perform the user authentication in uncontrolled environments – when a users naturally their phone.

There is indeed a significant prior art in authentication system implementations using various techniques such as speech [?], computer vision and image [?], graphical passwords [?], gestures [?], biometric fingerprints [?]. In this paper, we do acknowledge the viewpoint that our approach can also be used as a complementary scheme to most of the existing techniques in the authentication application space.

CONCLUDING REMARKS

We developed a system that uses head-movement patterns of users for direct authentication to a wearable device. We developed a light-weight approach that infers head-movements of users in response to music and generates signatures that are unique to every user. Our technique specifically uses the dynamic time-warping (DTW) tool to generate a head-movement feature that contains the mean and standard deviation of the DTW score of each user, determined over a multiple-user data set. The matching is conducted based on a thresholding scheme. Through a 30 user based experiment based evaluation using accelerometer traces collected using our data collection app on Google Glass, we observed that the average true-acceptance rate of our approach is at 95.1$\%$ and the false acceptance rates (FAR) at 3.9$\%$. From our evaluation over 30 subjects' data we observed that a 10 sec and $K = 3$ yield the least EER of 3.97$\%$. We observed that the FAR can be reduced by increasing the music cue duration or by using more samples for feature extraction (use Top K samples) or by increasing the training data set. We implemented *Headbanger* on Google Glass and profiled the execution time of the key system modules and observed response time of about 4.47 sec for a 10 sec music cue duration and reduces to 2.44 sec for a 5 sec music cue duration. Our profiling indicated that the most compute intensive part of the app (of the order of few seconds) was the classifier that involved DTW computation. We believe that such high compute times can be reduced through optimized algorithms or through strategic techniques such as pipelining the data input and execution process. The multi-user data sets were validated and verified during the course of our evaluations and will be released for public use in near future.