Activity-aware ECG-based Patient Authentication for Remote Health Monitoring

Janani Sriram
Department of Computer Science
Dartmouth College, Hanover, NH, USA
janani.sriram@dartmouth.edu

Minho Shin
Institute for Security, Technology, and Society
Dartmouth College, Hanover, NH, USA
mhshin@cs.dartmouth.edu

Tanzeem Choudhury
Department of Computer Science
Dartmouth College, Hanover, NH, USA
tanzeem.choudhury@dartmouth.edu

David Kotz
Institute for Security, Technology, and Society
Dartmouth College, Hanover, NH, USA
kotz@cs.dartmouth.edu

ABSTRACT

Mobile medical sensors promise to provide an efficient, accurate, and economic way to monitor patients' health outside the hospital. Patient authentication is a necessary security requirement in remote health monitoring scenarios. The monitoring system needs to make sure that the data is coming from the right person before any medical or financial decisions are made based on the data. Credential-based authentication methods (e.g., passwords, certificates) are not well-suited for remote healthcare as patients could hand over credentials to someone else. Furthermore, one-time authentication using credentials or trait-based biometrics (e.g., face, fingerprints, iris) do not cover the entire monitoring period and may lead to unauthorized post-authentication use. Recent studies have shown that the human electrocardiogram (ECG) exhibits unique patterns that can be used to discriminate individuals. However, perturbation of the ECG signal due to physical activity is a major obstacle in applying the technology in real-world situations. In this paper, we present a novel ECG and accelerometer-based system that can authenticate individuals in an ongoing manner under various activity conditions. We describe the probabilistic authentication system we have developed and present experimental results from 17 individuals.

[1]The research presented was supported by grants from Intel Corporation, by the U.S. Department of Commerce Award Number 60NANB6D6130, and by the U.S. Department of Homeland Security under Grant Award Number 2006-CS-001-000001. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of the sponsors.

ICMI-MLMI'09, November 2–4, 2009, Cambridge, MA, USA.

Categories and Subject Descriptors

K.6.5 [**Management of Computing and Information Systems**]: Security and Protection—*Authentication*; I.5.4 [**Pattern Recognition**]: Applications—*Signal Processing*

General Terms

Experimentation, Security

Keywords

Biometrics, ECG, Mobile Computing, Security

1. INTRODUCTION

As traditional healthcare systems struggle to cope with increasing demand and rising costs, many developed countries have been seeking new models of healthcare. Mobile computing and medical sensor technology offer a new paradigm for healthcare, namely *remote healthcare* [25, 19], that can reduce the cost [12] and improve the quality of healthcare services. Applications include long-term care for patients with chronic disease [22, 20], assistive technology for the elderly [21], risk management for people in rehabilitation [23], lifestyle coaching for people seeking to change unhealthy behavior [8], and fitness monitoring of athletes [18].

A remote health system relies on the ability of the system to correctly authenticate a patient, i.e., recognize whether the sensor data belongs to the right person, so that healthcare professionals can provide appropriate health services. Incorrect financial and medical decision could be made if someone else wears the patient's sensors, with or without the patient's permission. Existing authentication methods are not adequate for remote healthcare scenarios. If the authentication is based on passwords, smart cards, or secret keys, any person who has access to the credential can impersonate the patient.

Unlike traditional authentication methods, *biometric* authentication can discriminate between individuals based on their non-transferable traits such as face, voice, iris scan, or fingerprints. An authentication system usually operates in one of two different modes– *identification* and *verification*. During identification the

system selects the most probable candidate among a set of possible candidates (a multi-class classification problem). In verification mode the system decides whether the claimed identity is true or false (a binary classification problem). A one-time authentication has the risk of unauthorized post-recognition use, since someone else can wear the sensors after a successful authentication. Thus, it is desirable to have on-going authentication systems that confirms or rejects patient identity continuously.

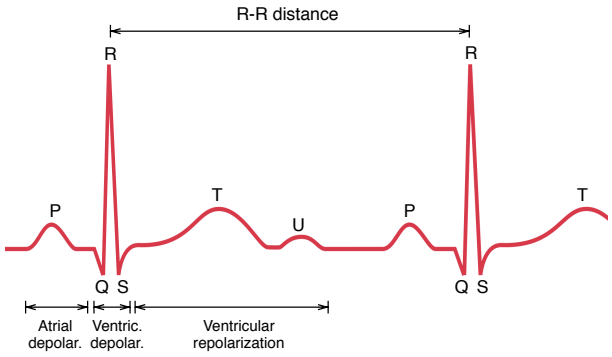


Figure 1: A typical normal ECG pulse consists of a P wave, which is associated with the contraction of the atria; a QRS complex, which is associated with the contraction of the ventricles; and T/U waves, which are associated with the repolarization of the ventricles.

The human ECG, an electrical signal that is associated with the electrical activity of the heart (Figure 1), offers several benefits as a biometric: it is universal, continuous, and difficult to falsify. The ECG signal from different individuals conforms to a fundamental morphology but also exhibits several personalized traits, such as relative timing of the various peaks, beat geometry, and responses to stress and activity. There are several factors that influence ECG signal: physiology and geometry of the heart, body build, gender, and age, which cause difficulties in accurate diagnosis [13, 15]. This inter-individual variability is potentially beneficial for discriminating individuals [36].

On the other hand, slow-changing factors, such as body habitus and age, can introduce long-term intra-subject variability, preventing permanent and unique identification of a person. Additionally, physiological responses to stimuli (such as stress and activity), and signal artifacts due to movement can result in large variability in a given individual's ECG signal. Intra-subject variability may also be caused by electrode placement [1], pharmaceutical drugs [35], and physical activities [32, 14]. Therefore, if ECG information is to be used for authentication in real situations it is necessary to account for the sources of significant intra-individual variability.

In this paper, we introduce an *activity-aware* biometric authentication system that combines ECG information with accelerometer data to handle the variability that arises from physical activity. Specifically,

- We develop a probabilistic ECG-based biometric authentication scheme that confirms a patient's identity in an ongoing manner across various activities.
- We implement a mobile prototype portable device using wearable ECG sensors.
- We demonstrate the feasibility of our patient-authentication scheme with 17 human subjects.

2. RELATED WORK

The validity of using the ECG in biometric recognition has been demonstrated in various related works, but most of the prior research considers ECG data collected from subjects at rest. For ongoing recognition, we would like to explore the impact of activity on the performance of the ECG based biometric system. Furthermore, we are interested in using a simple wearable ECG acquisition system and demonstrate the practical feasibility of the approach using a mobile prototype.

There are primarily two classes of feature extraction approaches for the ECG that are seen in related work - *fiducial* and *non-fiducial*. Fiducial points are points of interest within the heartbeat, such as local maxima or minima. Typical approaches extract different time and amplitude features using these points of reference [4, 16, 30]. Non-fiducial approaches aim to extract discriminative information from the ECG trace without having to localize fiducial markers. Thus, the fiducial features capture information local to a single heartbeat and the non-fiducial features capture global patterns (not restricted to a heart beat) in the ECG trace.

Biel et al. [4] used 12-lead ECG features to identify 20 subjects at rest. They use proprietary equipment to extract automatically extract 30 different fiducial features from each of the 12 leads, which is expensive to compute and more than is likely needed for biometric identification or verification. Furthermore, the 12-lead ECG system requires meticulous placement of 10 electrodes on each person, which can impede its usability.

Shen et al. [30] used 7 fiducial ECG-features from the most invariant part of the heartbeat: the QRS complex and T-wave. They then classify the features with a combination of template matching and decision-based neural network classifiers. Using one-lead ECG data, they achieved an identification accuracy of 100% for 20 individuals selected from the MIT/BIH database. The study demonstrates the potential of the ECG in biometric identification but does not investigate the impact of physical activity.

Israel et al. [16] developed an identification method based on fiducial features and studied the impact of seven states of mental stress on 29 subjects. Their experiments show that normalized temporal distances between fiducial points of ECG signal are invariant to anxiety state and can be used for biometric recognition under different stress levels. However, they do not consider the impact of physical activities.

The studies mentioned above and many other ECG-based identification and verification methods [2, 17] use the fiducial points from different parts of the ECG signal to extract features. However, detection of fiducial points is error-prone since there is no universally acknowledged rule for defining where the wave boundaries or fiducial points, lie [27].

Recent studies [7, 9, 27, 34] show that non-fiducial approaches can also successfully identify individuals from their ECG signal. Plataniotis et al. [27] and Wang et al. [34] have proposed an approach that uses autocorrelation analysis (AC) coupled with Discrete Cosine Transform (DCT) and does not require segmentation of the ECG trace into heart beats. On the other hand, Chiu et al. [9] used a discrete wavelet transform of ECG signal for feature extraction and Chan et al. [7] used a distance measure based on a wavelet transform. These studies also fail to consider the impact of activity that contributes to physiological variation (such as increased subject heart rate) and noise within the spectrum of the ECG signal.

The motion of the person can generate noise in the ECG signal because of unstable contact of an electrode with skin. Tong et al. [33] and Raya and Sison [28] each proposed a noise-cancelling technique to compensate for motion artifacts in ECG data, by using additional sensors: an accelerometer and an anisotropic magnetore-

sistive (AMR) sensor. In this paper, we have attempted to use the accelerometer information to model the variability rather than to correct variability with potential for loss of discriminatory information. During feature extraction we focus our efforts on the more robust part of the ECG signal - the QRS complex.

Even when accurate ECG measurements are available, the characteristics of the ECG signal exhibit physiological changes in response to activity [32, 14]. A few studies [17, 11] chose to complement the ECG signal with other biometric or non-biometric sensors. For example, Israel et al. [17] combined face recognition and an ECG signal to overcome limitations of the ECG signal as a biometrics. Damousis et al. [11] takes a broader approach to build a reliable authentication and monitoring system using trait-based biometrics (such as face or voice), physiology-based biometrics (such as ECG or electroencephalogram), and behavioral data (such as gait). In this paper, we handle activity-induced ECG variation by extracting a set of accelerometer features that characterize different physical activities along with fiducial and non-fiducial ECG features. We believe that such a multimodal feature set can provide the classifier with the necessary auxiliary information to perform accurate authentication. Table 1 summarises the comparison of the above-noted papers with our scheme. It should be noted that we cannot realistically compare our accuracy with different schemes due to several interpretation issues such the number of test samples per subject and the data collection methodology.

Reference	Feature	Method	Subjects	Activity	Acc
Biel [4]	Fiducial	PCA	20	No	100%
Shen [30]	Fiducial	Template Matching+DBNN	20	No	100%
Israel [16]	Fiducial	LDA	29	No	98%
Wang [34]	Both	KNN+LDA	13	No	96%
Chiu [9]	Non-Fiducial	Wavelet Distance+LDA	35	No	100%
Chan [7]	Non-Fiducial	wavelet DM	50	No	89%
Ours	Both	KNN+Bayesian	17	Yes	88%

Table 1: Comparison of related work with our scheme. The accuracy values represent the percentage of subjects who are correctly identified across a majority of their test samples.

3. METHODOLOGY

In this section, we describe the ECG plus accelerometer-based biometric authentication system we have developed and outline the preprocessing, feature extraction, and classification steps.

3.1 Sensors

As the base sensor board, we use the SHIMMER platform developed by Intel Digital Health Advanced Technology Group. The SHIMMER[1] is a compact sensing platform with an integrated 3-axis accelerometer. SHIMMER runs the TinyOS operating system and integrates (via custom cabling) to a commercially available Polar WearLink Plus ECG chest strap[2] . We ran our initial data collection experiments using the BioMOBIUS software. Sensor data from the SHIMMER's triaxial accelerometer and the ECG add-on board was sampled at 100Hz and transmitted via Bluetooth to BioMOBIUS which saves the data to a file.

3.2 Preprocessing and Feature Extraction

We segment the ECG and accelerometer traces into 400-sample windows (approximately 4 seconds of data) to obtain the feature windows $w_1, w_2, \ldots, w_n$. We use $w_i(j)$ to denote the j^{th} sample of feature window w_i. We chose the size of the feature window so that multiple heart beats are present within a given window (at least 4 in our case). We attempt to reduce the impact of misclassified R-peaks by averaging beat features within the window rather than relying on features extracted from a single beat. Although, this method introduces a time lag of 4s, we can reasonably assume that change of activity or person takes longer than 4s. We use the notation w_i^a and w_i^e to denote the accelerometer data and ECG data respectively within the i^{th} feature window. Any raw ECG trace collected using non-invasive surface electrodes usually has several artifacts, notably a low frequency baseline drift due to respiratory effects, electrode contact noise, and motion artifacts. Typically this noise is removed by high-pass or moving-average filtering techniques [28, 24, 10]. Since we collect our datasets during exercise, including durations of high-intensity activity, the ECG trace was also corrupted with the more troublesome motion artifact noise whose spectrum overlaps the ECG band (see Figure 2). The corresponding signal distortions cannot be easily eliminated by filtering.

We perform baseline correction before non-fiducial feature extraction. We employ an adaptive, beat-based linear interpolation approach to estimate the baseline from the line joining the q-minima. The estimated baseline is then subtracted to align all beats within a window. Baseline correction introduces some sharp discontinuities within the window, so before non-fiducial feature extraction we employ a high pass filter with coefficients adapted from [26].

We extract a combination of the two types of ECG features, fiducial and non-fiducial, from windows of the pre-processed signal. The final feature vector is shown in Table 2. Beat normalization ensures that the effects of beat-to-beat amplitude variations stay minimized. The normalized beats within a window have different periods since the adaptive segmentation in step 4 of Algorithm 1

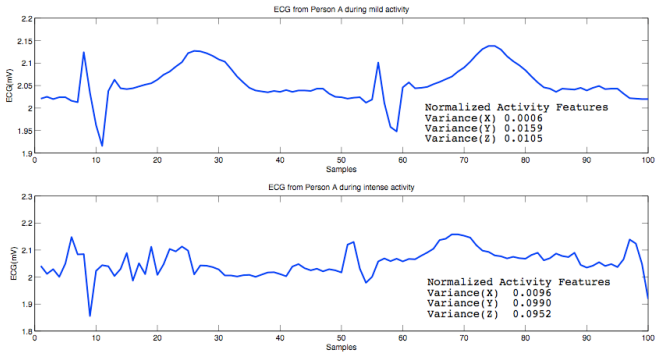


Figure 2: The ECG of a healthy subject at the beginning of exercise activity (above) and signal distortions due to motion artifact introduced as the subject proceeds to perform intense exercise activity (below).

	Feature
F1-F20	First 20 Normalized AC Coefficients
F21-F41	Spline interpolant
F42	Normalized Slope $\frac{QR}{QR+RS}$
F43	Normalized Slope $\frac{RS}{QR+RS}$
F44	R-R interval
F45-F50	Accelerometer X,Y,Z means and variances

Table 2: Feature Vector representing the Biometric Profile of an individual

[1] http://shimmer-research.com/
[2] http://polarusa.com/

is based on the distance between the R-peaks of the beats on either side of the current one. The underlying beat morphology is obtained from the mean spline (cubic) interpolant (F21–F41) of the beats within the feature window.

Fiducial analysis: Our procedure for beat segmentation and fiducial feature extraction is described in Algorithm 1. In addition, we

Algorithm 1 Detects the set of beats $\mathcal{B}$ and the sets of QRS markers, q,r,s, using an existing QRS detector [10]

1: The detected R-peaks are denoted by a set r of sample index-amplitude pairs as $(r_j, w_i^e(r_j))$, for the j^{th} beat of the i^{th} window of the ECG trace.
2: **for all** r_j in r **do**
3: Search downhill from each R-peak to locate the Q and S minima as $(q_j, w_i^e(q_j))$ and $(s_j, w_i^e(s_j))$. (The normal width of the QRS peak is known to be 100 ± 20ms [10]. We incorporate this fact by searching between ± 6 samples of the detected R-peaks.)
4: Align beat along the detected R-peaks by extracting a sequence of samples of size $\min((r_{j+1} - r_j), (r_j - r_{j-1}))$ so that the R-peaks are centered within the extracted beat segments. Discard if current beat $\mathcal{B}_j$ is incomplete and cannot be centered.
5: Normalize the beat by clamping the R-peaks to 1 and the Q-minima to 0.
6: **end for**
7: Discard beats corresponding to poorly detected fiducials that do not contain Q and S minima within the assumed search interval.
8: **return** $\{\forall j\ \mathcal{B}_j, (q_j, w_i^e(q_j)), (r_j, w_i^e(r_j)), (s_j, w_i^e(s_j))\}$

extract the mean R-R peak distances (F44) and the slopes QR(F42) and RS(F43). The RR-interval is normalized using the equation $\frac{x-l}{u-l}$ where x represents the current value and u and l represent bounds on the RR-interval (25 to 300 samples representing a heart rate range of 20 to 240 beats per minute). The slope features are normalized using the sum of the two slopes.

At the end of this stage we have a set of features that capture the underlying beat geometry and the activity induced variations. In particular, feature F44 models the intra-individual variations caused by physiological response to activity in the form of increased heart rate. Our set of features is a subset of those found in literature [4, 16, 30] with improved normalization and a novel adaptive beat-segmentation approach based on activity induced heart-rate changes.
Non-fiducial analysis: The presence of noise in the signal often leads to errors in beat segmentation of the ECG trace. So we complement the feature vector with a set of *non-fiducial* features that are less sensitive to the inaccuracies in beat segmentation [27, 7, 9]. The autocorrelation function of a signal represents how well the waveform correlates with a time-shifted (lags or leads) version of itself, i.e., its periodicity. For a sequence, x, of n samples representing the i^{th} ECG window $x = w_i^e$, the autocorrelation function is defined as

$$R_{xx}(m) = \sum_{i=0}^{n-|m|-1} x(i)x(i+m) \quad (1)$$

where lags $m = 0, 1 \ldots MAXLEAD$ where $MAXLEAD \ll n$

We use the normalized autocorrelation coefficients of each w_i^e as described by Plataniotis et al. [27]:

$$\tilde{R}_{xx} = \frac{R_{xx}(i)}{R_{xx}(0)}, \text{where } i = 1, 2 \ldots MAXLEAD. \quad (2)$$

We use the first 20 normalized coefficients ($MAXLEAD$=20) for each ECG window (n=400) as our non-fiducial features (F1–F20).
Accelerometer: The triaxial accelerometer measures the acceleration along the x, y and z-axes. We first zero-mean the signal by subtracting the mean of the entire trace. Then for each axis, we compute the mean and variance of the window w_i^a . The classifiers use these features to discriminate between different activities.

3.3 Classification

The goal of classification is to identify a subject or to verify an identity claim from the sensor observations. We investigate the performance of two types of classifiers: K-Nearest Neighbor (KNN) and Bayesian network(BN).

3.3.1 K-Nearest Neighbor

Initially, we measure the benefit of incorporating activity information using simple KNN classifiers. We trained two KNN classifiers: an activity-aware classifier, which uses the multimodal feature vector (F1–F50), and an activity-unaware classifier which uses the unimodal feature vector (F1–F44). We combine estimated pairwise correlation distances for features F1–F20 and F21–F42 and Euclidean distances for all other features to obtain a modified KNN classifier. KNN typically uses Euclidean distance as its distance metric. The additional correlation distance metric represents the similarity in the shapes of two curves. We evaluated the performance of both the Euclidean-distance-based KNN and the modified-KNN (xKNN) classifiers.

3.3.2 Bayesian Network

Our hypothesis is that explicit modeling of activity states will lead to better recognition performance. We developed two Bayesian network classifiers to allow us to evaluate that hypothesis (shown in Figure 3).

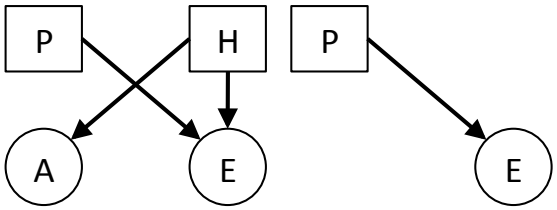
Suppose we have N persons whose identity P is given by the labels $p = p_1, p_2, \ldots, p_N$. The biometric profile for a person p_i is a set of m ECG feature vectors $e = \langle e_1, e_2, \ldots e_m \rangle$ and the corresponding activity feature vectors $a = \langle a_1, a_2, \ldots a_m \rangle$. We assume that the ECG features are normally distributed, i.e., $P(E \mid P, A) \sim N(\mu, \Sigma)$, and depends on the person and the activity being performed. The problem of classification is now reduced to that of estimating the parameters of the conditional distributions of Equation (3). We discretize the accelerometer features,A, into distinct activity levels H and obtain the Bayesian Network (BN) shown in Figure 3(a). The joint distribution of the BN is defined as

$$P(P, H, A, E) = P(E \mid P, H)P(A \mid H)P(H)P(P) \quad (3)$$

During training, we assume that the BN is fully observed.

The activity levels H were obtained in two ways: from manual annotations and via unsupervised clustering. Manual annotations include three activity levels: still, low-intensity, and high-intensity. We also tested three types of unsupervised clustering techniques: K-means clustering with the Euclidean distance metric, K-means using city block distance and Gaussian Mixture Model (GMM) with diagonal covariance matrices. Ultimately, we chose the GMM which allows soft cluster boundaries and resulted in better performance. Since the manual annotations contained significant human error, there was no reliable ground truth for comparing the activity clustering. So we visually compared the inferred cluster IDs with our manual annotations and found that the inferences were close except for occasional discontinuities (see Figure 4).

We used the Bayes Net Toolbox [3] for training and inference. During testing both nodes H and P were hidden. We are interested in estimating the probability distribution of the hidden variable P,



(a) The Activity-Aware Bayesian Network.

(b) The Activity-Unaware Bayesian Network

Figure 3: Nodes: Person ID $p = 1, \ldots, N$ for N subjects. The feature nodes A and E are 6- and 44-dimensional Gaussians respectively. Node H represents the discrete activity labels.

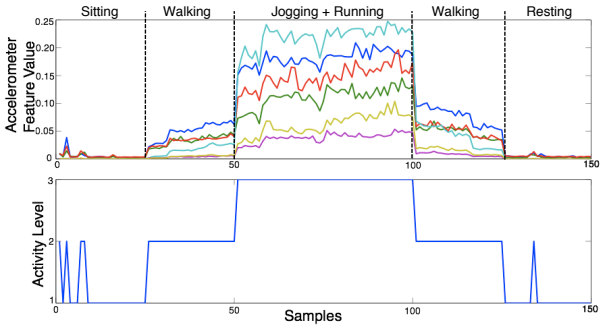


Figure 4: Unsupervised Activity Clustering. Values for the six activity feature are shown above and the corresponding discrete activity levels are shown below.

conditioned on the accelerometer and ECG features $A = a$ and $E = e$ respectively, i.e.,

$$P(P \mid A = a, E = e) = \sum_{j=1}^{|h|} P(P, H = h_j \mid A = a, E = e) \tag{4}$$

The predicted person label p_{pred} is the one that maximizes the posterior probability.

$$p_{pred} = \underset{P}{\operatorname{argmax}} \sum_{j=1}^{|h|} P(P, H = h_j \mid A = a, E = e) \tag{5}$$

To evaluate the usefulness of the activity features, we also test an activity-unaware the Bayesian classifier shown in Figure 3(b), which uses only the set of ECG features.

Identification: During identification, the system selects the person with highest marginal probability.
Verification: During verification we are only interested in accepting or rejecting a claimed identity. One possible approach is to compare the probabilities estimated during identification against a threshold to obtain verification decisions.

Another approach is to perform binary classification using two class labels, $p = \{\text{legitimate,imposter}\}$ – which leads to a more compact probabilistic model and maybe preferable when there are a large number individuals, especially by mobile inference systems.

To build a binary verification model, we need a representative set of imposters for each individual. Selection of imposters has been the subject of much research in related studies of person verification using speech, signatures and other modalities that exhibit significant intra-subject variability [5, 29, 31]. We adopt two principles from this body of work: the selection of subject-specific imposters and the pooling of samples from multiple imposters. We describe our verification algorithm in Algorithm 2, in which we denote the training data for persons $p = \{p_1, p_2 \ldots p_N\}$ to be $\mathcal{D}(p_1), \mathcal{D}(p_2) \ldots \mathcal{D}(p_N)$.

Algorithm 2 Training the pooled imposter model for Verification

1: **for all** claimants $p_i \epsilon\, p = \{p_1, p_2 \ldots p_N\}$ **do**
2: select a set of k most confused subjects as imposters $\hat{p} = \{\hat{p_1}, \hat{p_2}, \ldots \hat{p_k}\}$ such that $\hat{p_j} \epsilon\, p$ and $\hat{p_j} \neq p_i$ for any j
3: obtain the legitimate claimant dataset $\mathcal{T}$ as $\mathcal{D}(p_i)$
4: **for** $j = 1$ to k **do**
5: add $\mathcal{D}(\hat{p_j})$ the pooled imposter dataset $\mathcal{F}$
6: **end for**
7: Randomly sample from the pooled imposter dataset such that $\mathcal{T}$ and $\mathcal{F}$ are of equal sizes.
8: Estimate model parameters for p_i, training data $\bar{\mathcal{D}} = [\mathcal{T}, \mathcal{F}]$
9: **end for**

Finally, the system uses the verification score, given as the ratio of true and imposter model likelihoods, to make a decision about a claimed identity p_i:

$$\frac{P(p_i \text{ is a legitimate claimant})}{P(p_i \text{ is an imposter})} \tag{6}$$

Since both probabilities are obtained from the claimant's model the proportional score simply serves to exaggerate differences in their probabilities for comparison against a threshold.

The potential to misclassify subjects who are not represented in the imposter pool is one of the drawbacks of the proposed verification model. But the compact representation and lower computational cost makes it a potentially appealing option. When we tested a dataset of 20 samples on the identification model running on an Intel Core2 Duo machine, it took approximately 0.6 seconds to predict all class labels. A verification decision on the same dataset took around 0.2s.
Combining Multiple Predictions: We refer to the individual prediction for each feature window as window identification/verification. A sequence of window identification/verification decisions are used to make person identification/verification decisions by majority voting. During identification, an entire test sequence is classified as belonging to person p_i if the majority of the predictions are for p_i. During verification, an entire test sequence is verified for a certain claim p_i if the majority of the predictions are verified as legitimate. For ongoing verification, instead of majority voting over the entire dataset, we evaluate the performance of the classifier as follows,

1. Divide test dataset in chunks of 10 feature windows each.
2. For each chunk if the majority of windows is classified as legitimate, the claimant is considered legitimate. If not the claimant is an imposter.

4. EXPERIMENTAL RESULTS

To test the feasibility of our approach, we collected data from 17 volunteers under different activity conditions and across different days. To make it easier to compare with related approaches, we present the *identification* performance of the classifiers in addition to the verification results. We also describe the wearable prototype system that we built in Section 4.3.

4.1 Data Collection

We collected sensor data from normal healthy subjects for the duration of a workout session. Subjects were asked to exercise on the treadmill for 12–15 minutes (training dataset DT) or 5–7 minutes (test dataset DX). We collected test and training data on different days. The subjects were told to pace themselves and slowly work up to a jog; after at least 2 minutes of brisk jogging they were asked to slow down gradually to a halt. The data is manually labeled with annotations about the subjects pace and duration of exercise. We collected sitting (DS) and recovery (DR) data immediately before and after the training work-out. To construct the training dataset (DT, DS, DR) we selected approximately 10000 samples from each of the following activity annotations – sit, walk, run, endure run and recover. These are grouped into either 3 activity level annotations: sit+recover (labeled: still),walk (labeled: low intensity), run+endure run (labeled: high intensity) or 2 activity levels annotations: sit+recover (labeled: still) and walk+run+endure run (labeled: high+low intensity) for the supervised models. Some subjects had fewer than 10000 samples per annotation due to noise. In aggregate the training dataset size for each subject ranged between 40000-50000 samples. We tested our classifiers using samples from DX.

4.2 Identification and Verification Results

In order to compare with related approaches, we first evaluate the performance of the classifiers on data taken from subjects at rest (DS). The results are presented in Table 3 and are comparable to other existing approaches.

	KNN	xKNN	BN
Precision	0.96	0.97	0.97
Recall	0.95	0.97	0.97

Table 3: Identification performance for the sitting session (DS). We randomly select 20 windows as the test dataset and the remaining data as the training dataset.

Next we tested the performances of the different activity-aware and activity-unaware classifiers on the test dataset DX. Generally, the activity-aware classifiers outperformed the activity-unaware classifiers (as shown in Table 4) by being able to explain the intra-subject variability seen in the ECG signal. Among the activity-aware classifiers, the KNN classifiers do not explicitly model the effects of different activity levels. Nonetheless, the simple inclusion of the additional accelerometer modality is clearly useful, as seen by the improvement in performance. We use k=1 (one neighbor) based on our results from cross-validation. The BN classifier makes better use of the accelerometer data by explicitly modeling the activities, which leads to slightly better performance numbers even with the manual activity annotations.

Manual annotations are inconvenient, unreliable, subjective and unsuitable for fine-grained activity clustering (beyond a small number of levels). We used activity labels derived from unsupervised activity clustering using Gaussian mixture models. Table 5 shows the performance of the Bayesian classifier based on unsupervised activity clustering. The case where $|H|$=1 corresponds to the activity-unaware classifier.

Table 6 shows the performance of verification with varying number of imposters. Verification decisions are made according to Equation(5). We can view the false positive rate (FPR) as an indicator of the security of the system. As would be expected, FPR is reduced as more imposters were added to the pool. It should be noted that we evaluate the entire test dataset from all subjects against every combination of claimant and imposter to obtain the results shown in Table 6. When the number of imposters is large, the random sampling did not include sufficient samples from each activity and imposter combination. This resulted in an increase in false acceptances possibly due to an ill-constructed pooled dataset. With too few imposters the pooled dataset does not have enough information to begin with.

	Activity-Aware				Activity-Unaware		
	KNN	xKNN	BN		KNN	xKNN	BN
			$\|h\|=2$	$\|h\|=3$			
Precision	0.8243	0.8278	0.8488	0.8252	0.7855	0.7677	0.8139
Recall	0.8039	0.7925	0.8326	0.8174	0.8035	0.7987	0.8140

Table 4: Identification performance of the activity-aware classifiers against the activity-unaware classifiers. The activity-aware KNN classifiers use a concatenated feature vector of activity and ECG features. The activity-aware BN is provided supervised activity labels derived from manual annotations. An improvement is apparent even with just two activity levels.

	Window Identification		Person Identification
No. of Activity Levels	True Positive Rate	False Positive Rate	Accuracy
1	81.39	5.65	14/17
2	77.59	5.63	15/17
3	78.67	5.65	15/17
4	79.34	5.57	15/17
5	78.03	5.52	15/17

Table 5: Identification performance for the Bayesian network classifier using unsupervised activity clustering for varying number of activity clusters. A test session consists of a sequence of ECG feature windows extracted from dataset DT. The window identification treats each feature vector as a separate test data point. The person identification decision combines results from all the windows and selects the person predicted by majority of the windows.

	Window Verification		Person Verification
No. of Imposters	True Positive Rate	False Positive Rate	Accuracy
3	81.68	12.72	16/17
7	81.14	11.72	16/17
8	79.31	11.13	16/17
11	81.7	12.19	17/17
15	79.85	12.06	16/17

Table 6: Verification performance for different sizes of imposter pools. The test dataset consists of all windows from all persons tested for every possible claimant. A person is considered correctly verified if a majority of his samples are verified as legitimate, i.e., TPR>0.5.

Figure 5 shows the ROC curve for the verification model. Depending on the nature of application, we can choose to optimize for fewer false positives or more true positives by selecting different operating points on the curve . To help reduce the number of overall false rejection, we agregate the window verification decision. As described in Section 3.3.2, we combine 10 window predictions in an ongoing manner in our experiments. We use predictions from the 8 person imposter model. The results in Table 7 show, for each person, how often his identity claims are correctly accepted (legitimate claimant) and how often the claims made by the most confused imposter are accepted. The most confused imposter test provides the worst case verification numbers. Fewer imposter claims are incorrectly accepted for the other imposters in the dataset.

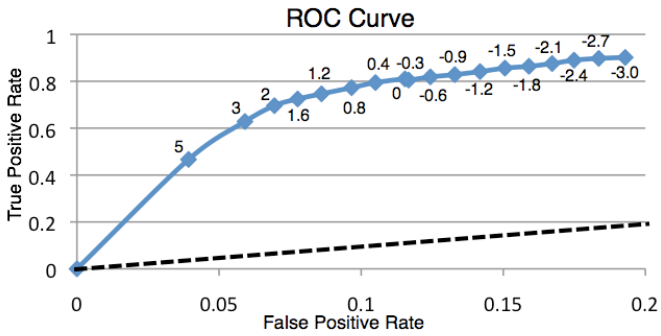


Figure 5: ROC curve for the verification model (8 imposters). The thresholds used are shown in the graph. We can see that with higher thresholds the system rejects too many legitimate users and with lower thresholds too many imposters are accepted. The dotted line shows y=x.

	Legitimate Claimant	Imposter
p1	100.00	0.00
p2	88.24	75.00
p3	90.91	0.00
p4	100.00	0.00
p5	100.00	0.00
p6	100.00	27.27
p7	81.82	0.00
p8	100.00	70.00
p9	85.71	12.50
p10	80.00	0.00
p11	81.82	0.13
p12	93.75	83.3
p13	91.67	0.00
p14	100.00	0.00
p15	77.78	0.00
p16	100.00	0.00
p17	60.00	0.00

Table 7: Acceptance rates for person verification (8 imposters) by aggregating window verification decisions. A person verification decision is made based on the most number of verified samples within a chunk. The test dataset consists of data from the legitimate claimant and imposter (most confused).

4.3 Prototype Application

To test the feasibility of our biometric verification approach, we implemented a prototype application as shown in Figure 6. Our primary design goal for the prototype was to set up a simple and reliable architecture to relay data from the sensor to an authentication server that runs the pattern recognition algorithms. We implemented a NesC application running on the SHIMMER to send ASCII packets of data from the onboard accelerometer and the add-on ECG sensor board over Bluetooth [6]. A thin client application running on a mobile phone (Nokia N95) wirelessly receives sensor data over Bluetooth and forwards it to the remote authentication server over an IEEE802.11g (Wi-Fi) link. Our protocol forwards chunks of 4000 samples so that data is sent to the authentication server every 40s. We implemented the analysis algorithms in MATLAB. The authentication server periodically analyzed the sensor data and logged the results. Although this approach works, significant development and evaluation still needs to be done to make this application practical and usable.

5. SUMMARY

In this paper we introduced a novel multimodal biometric authentication system based on wearable human electrocardiogram

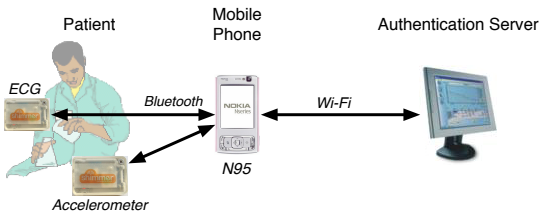


Figure 6: Prototype Architecture

(ECG) and accelerometer sensors. We demonstrated, on data collected from 17 subjects, that activity-aware authentication systems can effectively deal with the ECG variability induced by physical activities performed in the real world. We believe the approached outline in the paper could facilitate ongoing authentication without requiring frequent and active participation from the user. In our future work, we plan include a broader range of activities and potentially more sensors. For instance, a galvanic skin sensor may be used to classify stressful activities. We also plan to develop the mobile verification platform further and perform verification locally on the mobile device.

6. REFERENCES

[1] J. Aganauskiene, L. Sornmo, R. Atarius, and C. Blomstrom-Lundqvist. Reproducibility of the signal-averaged electrocardiogram using individual lead analysis. *European Heart Journal*, 16(9):1244–1254, 1995.

[2] F. Agrafioti and D. Hatzinakos. Fusion of ECG sources for human identification. In *Proceedings of the 3rd International Symposium on Communications, Control and Signal Processing (ISCCSP)*, pages 1542–1547, March 2008. DOI: 10.1109/ISCCSP.2008.4537472.

[3] Bayes Net Toolbox (BNT). As viewed July 2009. http://www.cs.ubc.ca/~murphyk/Software/BNT/bnt.html.

[4] L. Biel, O. Pettersson, L. Philipson, and P. Wide. ECG analysis: a new approach in human identification. In *Proceedings of the 16th IEEE Instrumentation and Measurement Technology Conference*, volume 1, pages 557–561, 1999. DOI: 10.1109/IMTC.1999.776813.

[5] Frédéric Bimbot, Jean-François Bonastre, Corinne Fredouille, Guillaume Gravier, Ivan Magrin-Chagnolleau, Sylvain Meignier, Teva Merlin, Javier Ortega-García, Dijana Petrovska-Delacrétaz, and Douglas A. Reynolds. A tutorial on text-independent speaker verification. *EURASIP Journal of Applied Signal Processing*, 2004:430–451, 2004. DOI: dx.doi.org/10.1155/S1110865704310024.

[6] Bluetooth Special Interest Group. Specification of the Bluetooth System: Core V2.1 + EDR, July 2007.

[7] A. D. C. Chan, M. M. Hamdy, A. Badre, and V. Badee. Wavelet distance measure for person identification using electrocardiograms. *IEEE Transactions on Instrumentation and Measurement*, 57(2):248–253, February 2008. DOI: 10.1109/TIM.2007.909996.

[8] C. V. Chan and D. R. Kaufman. Mobile phones as mediators of health behavior change in cardiovascular disease in developing countries. *Studies in Health Technology and Informatics*, 143:453–458, 2009. DOI: 10.3233/978-1-58603-979-0-453.

[9] C. Chiu, C. Chuang, and C. Hsu. A novel personal identity verification approach using a discrete wavelet transform of the ECG signal. In *Proceedings of the International Conference on Multimedia and Ubiquitous Engineering, MUE 2008*, pages 201–206, April 2008. DOI: 10.1109/MUE.2008.67.

[10] G. D. Clifford, F. Azuaje, and P. McSharry. *Advanced methods and tools for ECG data analysis*. Artech House, Inc., 2006.

[11] I. G. Damousis, D. Tzovaras, and E. Bekiaris. Unobtrusive multimodal biometric authentication: The HUMABIO project concept. *EURASIP Journal on Advances in Signal Processing*, pages 1–11, 2008. DOI: 10.1155/2008/265767.

[12] Susan L. Dimmick, Samuel G. Burgiss, Sherry Robbins, David Black, Bertha Jarnagin, and Mary Anders. Outcomes of an integrated telehealth network demonstration project. *Telemedicine Journal and e-Health*, 9(1):13–23, March 2003. DOI: 10.1109/MC.2008.133.

[13] I. Eisenstein, J. Edelstein, R. Sarma, M. Sanmarco, and R. H. Selvester. The electrocardiogram in obesity. *Journal of Electrocardiology*, 15(2):115–118, April 1982.

[14] N. J. Fortuin and J. L. Weiss. Exercise stress testing. *Journal of the American Heart Association*, 56(5):699–712, 1977.

[15] R. Hoekema, G. J. H. Uijen, and A. van Oosterom. Geometrical aspects of the interindividual variability of multilead ECG recordings. *IEEE Transactions on Biomedical Engineering*, 48(5):551–559, May 2001. DOI: 10.1109/10.918594.

[16] S. A. Israel, J. M. Irvine, A. Cheng, M. D. Wiederhold, and B. K. Wiederhold. ECG to identify individuals. *Pattern Recognition*, 38(1):133–142, January 2005. DOI: 10.1016/j.patcog.2004.05.014.

[17] S. A. Israel, W. T. Scruggs, W. J. Worek, and J. M. Irvine. Fusing face and ECG for personal identification. In *Proceedings of the 32nd Applied Imagery Pattern Recognition Workshop*, pages 226–231, October 2003. DOI: 10.1109/AIPR.2003.1284276.

[18] D. Jea, J. Liu, T. Schmid, and M. B Srivastava. Hassle free fitness monitoring. In *Proceedings of the 2nd International Workshop on Systems and Networking Support for Healthcare and Assisted Living Environments (HealthNet)*, June 2008. DOI: 10.1145/1515747.1515756.

[19] A. D. Jurik, Jonathan F. Bolus, A. C. Weaver, B. H. Calhoun, and T. N. Blalock. Mobile health monitoring through biotelemetry. In *Proceedings of the International Conference on Body Area Networks (BodyNets)*, April 2009.

[20] H. S. Kim and H. S. Jeong. A nurse short message service by cellular phone in type-2 diabetic patients for six months. *Journal of Clinical Nursing*, 16(6):1082–1087, June 2007. DOI: 10.1111/j.1365-2702.2007.01698.x.

[21] S. Lee, Y. Kim, G. Lee, B. Cho, and N. Lee. A remote behavioral monitoring system for elders living alone. In *Proceedings of the International Conference on Control, Automation and Systems (ICCAS)*, pages 2725–2730, October 2007. DOI: 10.1109/ICCAS.2007.4406830.

[22] A. G. Logan, W. J. McIsaac, A. Tisler, M. J. Irvine, A. Saunders, A. Dunai, C. A. Rizo, D. S. Feig, M. Hamill, M. Trudel, and J. A. Cafazzo. Mobile phone-based remote patient monitoring system for management of hypertension in diabetic patients. *American Journal of Hypertension*, 20(9):942–948, September 2007. DOI: 10.1016/j.amjhyper.2007.03.020.

[23] A. Lymberis. Smart wearables for remote health monitoring, from prevention to rehabilitation: current R&D, future challenges. In *Proceedings of the 4th International IEEE EMBS Special Topic Conference on the Information Technology Applications in Biomedicine*, pages 272–275, April 2003. DOI: 10.1109/ITAB.2003.1222530.

[24] M. Milanesi1, N. Martini, N. Vanello, V. Positano, M. F. Santarelli, and L. Landini. Independent component analysis applied to the removal of motion artifacts from electrocardiographic signals. *Medical and Biological Engineering and Computing*, 46:251–261, 2008. DOI: 10.1007/s11517-007-0293-8.

[25] N. Oliver and F. Flores-Mangas. HealthGear: a real-time wearable system for monitoring and analyzing physiological signals. In *Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks (BSN)*, pages 4–64, April 2006. DOI: 10.1109/BSN.2006.27.

[26] Jiapu Pan and Willis J. Tompkins. A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, 32(3):230–236, March 1985. DOI: 10.1109/TBME.1985.325532.

[27] K. N. Plataniotis, D. Hatzinakos, and J. K. M. Lee. ECG biometric recognition without fiducial detection. *Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference*, 19:1–6, August 2006. DOI: 10.1109/BCC.2006.4341628.

[28] M. A. D. Raya and L. G. Sison. Adaptive noise cancelling of motion artifact in stress ECG signals using accelerometer. In *Proceedings of the Second Joint EMBS/BMES Conference*, volume 2, pages 1756–1757, 2002. DOI: 10.1109/IEMBS.2002.1106637.

[29] A. E. Rosenberg and S. Parthasarathy. Speaker background models for connected digit password speaker verification. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 81–84, 1996. DOI: dx.doi.org/10.1109/ICASSP.1996.540295.

[30] T. W. Shen, W. J. Tompkins, and Y. H. Hu. One-lead ECG for identity verification. *Proceedings of the 24th Annual Conference on Engineering in Medicine and Biology and the Annual Fall Meeting of the Biomedical Engineering Society (EMBS/BMES)*, 1:62–63, 2002. DOI: 10.1109/IEMBS.2002.1134388.

[31] T. Sim, Sheng Zhang, R. Janakiraman, and S. Kumar. Continuous verification using multimodal biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):687–700, April 2007. DOI: 10.1109/TPAMI.2007.1010.

[32] M. L. Simoons and P. G. Hugenholtz. Gradual changes of ECG waveform during and after exercise in normal subjects. *Journal of the American Heart Association*, 52:570–577, 1975.

[33] D. A. Tong, K. A. Bartels, and K. S. Honeyager. Adaptive reduction of motion artifact in the electrocardiogram. In *Proceedings of the Second Joint EMBS/BMES Conference*, volume 2, pages 1403–1404, 2002. DOI: 10.1109/IEMBS.2002.1106451.

[34] Y. Wang, F. Agrafioti, D. Hatzinakos, and K. N. Plataniotis. Analysis of human electrocardiogram for biometric recognition. *EURASIP Journal on Advances in Signal Processing*, (1):1–6, 2008. DOI: 10.1155/2008/148658.

[35] M. Wolzt, L. Schmetterer, J. Kastner, K. Krejcy, G. Zanaschka, C. Unfried, and H. G. Eichler. Short-term drug effects on the signal-averaged electrocardiogram in healthy men: assessment of intra- and interindividual variability of spectral temporal mapping and time-domain analysis. *Journal of Pharmacology and Experimental Therapeutics*, 275(3):1375–1381, 1995. Online at http://jpet.aspetjournals.org/cgi/content/abstract/275/3/1375.

[36] G. Wuebbeler, R. Bousseljot, D. Kreiseler, M. Stavridis, and C. Elster. Human verification by heart beat signals, Working Group 8.42, Physikalisch-Technische Bundesanstalt(PTB), berlin, germany, 2004. Online at http://www.berlin.ptb.de/8/84/842/BIOMETRIE/842biometriee.html.