A Behavioral Biometrics Based Approach to Online Gender Classification

Nicolas Van Balen[1(✉)], Christopher T. Ball[1], and Haining Wang[2]

[1] College of William and Mary, Williamsburg, VA 23187, USA
njvanbalen@email.wm.edu
[2] University of Delaware, Newark, DE 19716, USA

Abstract. Gender is one of the essential characteristics of personal identity but is often misused by online impostors for malicious purposes. However, men and women differ in their natural aiming movements of a hand held object in two-dimensional space due to anthropometric, biomechanical, and perceptual-motor control differences between the genders. Exploiting these natural gender differences, this paper proposes a naturalistic approach for gender classification based on mouse biometrics. Although some previous research has been done on gender classification using behavioral biometrics, most of them focuses on keystroke dynamics and, more importantly, none of them provides a comprehensive guideline for which metrics (features) of movements are actually relevant to gender classification. In this paper, we present a method for choosing metrics based on empirical evidence of natural difference in the genders. In particular, we develop a novel gender classification model and evaluate the model's accuracy based on the data collected from a group of 94 users. Temporal, spatial, and accuracy metrics are recorded from kinematic and spatial analyses of 256 mouse movements performed by each user. A mouse signature for each user is created using least-squares regression weights determined by the influence movement target parameters (size of the target, horizontal and vertical distances moved). The efficacy of our model is validated through the use of binary logistic regressions.

1 Introduction

The popularity of online social networks, online forums, and various online dating sites has significantly increased the visibility of online users' personal information. However, these online sites also allow a great deal of anonymity in the sense that a user's identity is tied to the user's account but not personally to the user. This anonymity has been exploited by impostors, such as sexual predators, who lie about their gender or age for malicious purposes, while a victim user has little way of verifying that the provided information is valid. To date, very little has been done to address this problem of fake online personal identity. A strict registration policy, such as providing legal documents, is just not feasible for regulating this problem.

R. Deng et al. (Eds.): SecureComm 2016, LNICST 198, pp. 475–495, 2017.
DOI: 10.1007/978-3-319-59608-2_27

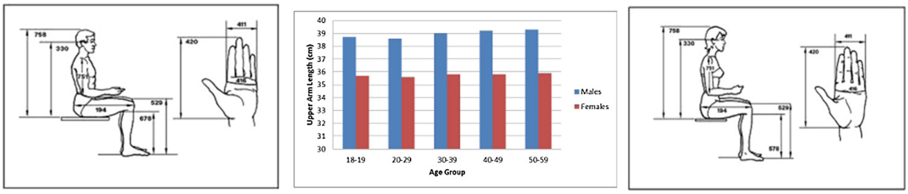

Fig. 1. Illustration of the major anatomical measurements relevant to using a computer mouse from a seated position. Graph of gender differences in upper limb length (data taken from Anthropometric Reference Data for Children and Adults: Unites States, 2007–2010; U.S. Department of Health and Human Services) [1].

One promising alternative involves the use of physical or behavioral biometrics, such as keystroke dynamics or mouse dynamics, to enhance user authentication. These biometrics are non-invasive and can be used actively as a confirmation step or passively through continuous re-authentication to determine the demographic characteristics of a user. However, previous soft biometric systems tend to take a very data driven approach based on simple aggregate measures (e.g., averages) of behavioral metrics. In this paper, we present a new naturalistic approach to using behavioral biometrics for verifying an online user's demographics. We will illustrate the advantages of this approach by applying mouse biometrics to discriminate a user's gender. Our approach takes advantage of intra-user variability in mouse movements, and has the potential to overcome generalizability issues when using mouse biometrics for user verification.

The proposed approach is mainly based on two important assumptions regarding naturally occurring mouse movements: (1) Gender differences naturally exist when performing two-dimensional aiming movements of a hand held device. The support for this assumption comes from a variety of basic and applied research domains, which include occupational health, physical therapy, public health, ergonomics, human anatomy, and perceptual-motor control theory. (2) The gender differences alluded to in the first assumption can be further elaborated by tracking the changes to naturally occurring mouse movements that are imposed by different target parameters. These target parameters are defined by the horizontal and vertical distances between the start and endpoint target locations, and by the size of the endpoint target. All three task parameters are known to affect aiming movements [11,25,28] while recent research in perceptual-motor control has highlighted that gender can also mediate these effects [4,23,24].

As a result of these two assumptions, this approach incorporates a much wider array of mouse movement metrics than those used in previous security applications of mouse biometrics. Consequently, the data analysis of these metrics required a different statistical approach from that used in traditional investigations of mouse biometrics. Twenty one different mouse movement metrics (temporal, spatial, and accuracy) were extracted from the movements recorded, and then each metric was expressed as a vector of four variables. The four variables correspond to the intercept and three unstandardized regression coefficients that are obtained from a multiple regression equation formulated to predict each metric using the three target parameters (vertical distance, horizontal distance, and target size). Binary logistic regressions were then employed to predict each participant's gender using an optimal subset of the multiple regression coefficients.

The proposed model was validated with mouse movement data collected from 94 participants (45 male and 49 female) who each performed 256 movement trials. Our user data collection has been filed and approved by the Institutional Review Board (IRB) to ensure participants are treated ethically. The model's accuracy was tested on both labeled and unlabeled data. The labeled data is used as a verification step to test our method's ability to accurately fit the model to the real data and identify a user that has uncommon mouse movement characteristics as an outlier, while the unlabeled data is used to test the ability to accurately classify a user who has not yet been sighted before. Based on the evaluation results in both labeled and unlabeled data, an analysis of the outliers' impact was further performed to test the impacts that outliers, i.e., those users with mouse movement characteristics greatly different from the average, would have on the model. The achieved maximum accuracy is 89.4% for the full labeled data set and 100% after removing outliers, while 72.4% for the unlabeled data set and 75.9% after removing outliers.

The remainder of the paper is structured as follows. Section 2 describes the logic behind the naturalistic approach, along with a summary of related work. Section 3 details the methodology used to collect data, filter data, and extract the metrics from the data to be used for gender classification. Section 4 presents the two analysis steps used in building the statistical models for predicting the gender of each participant. Section 5 reports the results of testing the statistical models. Section 6 reviews the findings and limitation of the study, as well as describing future directions for this naturalistic approach. Finally, Sect. 7 summarizes the paper.

2 Background

In this section, we first highlight the gender difference in anthropometrics, including its induced differences in movement behaviors and grip postures. We then present the background of using behavioral biometrics for user authentication.

2.1 Gender Difference in Anthropometrics

Men and women clearly differ in their physical dimensions as described by anthropometric data recorded in many countries for the purposes of monitoring public health and designing ergonomically sound work environments. Figure 1 illustrates the important anthropometric attributes of an individual working with a typical computer system. Maneuvering a computer mouse across a 2-Dimensional work space requires the complex coordination of the upper and lower arms in combination with the wrist and fingers. As shown in Fig. 1, the anthropometric data for the upper arm length (reported by the United States Health Department [1]) reveals large consistent gender differences in the physical dimensions of a key limb component for moving a mouse on a table top. Physical differences like these arguably underlie many of the movement and grip differences that will be described in the remainder of this section [17].

Moving a computer mouse is classified as an aiming movement by researchers in the field of motor behavior, and aiming movements are generally composed of consistent temporal and spatial characteristics. An aiming movement typically includes a ballistic component (single phase of acceleration followed by deceleration) that corresponds to the main movement of the hand into the general area of the target location. The ballistic component is followed by a sequence of sub-movements (multiple phases of acceleration and deceleration) that consist of small spatial corrections of the hand to reach the final target destination [20]. The field of motor behavior suggests that men and women differ in their aiming movements with men tending to move faster than women and with less accuracy [4,6,9,23,27]. It was also reported that the location of the target in relation to the hand being used affected the accuracy of movements made by men, but showed no significant effect on women's movements [23]. These results not only highlight gender differences in movement behavior again, but also stress the importance of incorporating target parameter effects when investigating these gender differences. Here the target parameters include target size, horizontal distance, and vertical distance.

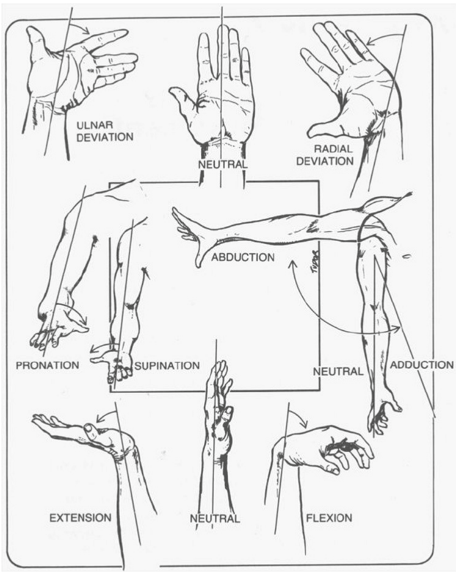

Fig. 2. Anatomical terms for motions of upper limb, wrist, and joints.

Research in physical therapy that has examined the effects of mouse use on wrist and arm pain in computer users has shown gender differences in hand and arm postures when performing movements with a mouse. A study on the finger postures of mouse users showed that men more frequently had a finger posture, in which the finger used for mouse clicking had a lifted finger posture where the middle portion of the finger was not in contact with the mouse [18]. Male participants in this study were also more likely to show an extended finger posture with a flexion angle of less than 15° when gripping the mouse (refer to Fig. 2 for an illustration of relevant movement terms). These different grip postures may not only affect mouse movement characteristics, but also influence mouse button presses that can also be an important component of mouse biometrics. Johnson et al. [16] found that women exerted more relative force on the mouse when gripping it, while Wahlstrom et al. [30] reported that women exerted more force on the mouse button while pressing it. Johnson and colleagues also revealed different wrist postures between men and women when moving the mouse with women showing higher wrist extensions, larger ulnar deviations (refer to Fig. 2), a larger range of motion in the wrist, and higher wrist velocities. A similar study by Yang and Cho [32] reported larger elbow flexion angles in men as well as different ulnar deviations, but in this study it was the men who exhibited the larger ulnar deviation angles. All of these different grip postures have the potential to affect mouse movement characteristics, including mouse button presses that can also be an important component of mouse biometrics. The results of these studies suggest that mouse biometrics should not only consider movement characteristics of aiming movements, but also consider movement characteristics unique to the physical manipulation of gripping a computer mouse.

2.2 Behavioral Biometrics

The use of biometrics is an attractive option for user authentication since it is inherently based on "who you are," and unlike other conventional methods cannot be lost, forgotten, or stolen. A large variety of user characteristics are used in biometric identification with some involving physiological recording, such as iris scanning, fingerprint scanning, facial recognition, and pulse recording [22][1]; and some involving behavioral recording, such as keystroke and mouse dynamics [26,31]. The behavioral biometric systems, however, have the distinct advantage of not requiring specialized hardware to record the user behaviors. Research interest in behavioral biometrics started in the 1990s with the study of keystroke dynamics [19] that eventually led to research involving keystroke dynamics combined with mouse dynamics [2].

Behavioral biometrics have been used in the past to predict the gender of a user, but these studies have primarily focused on keystroke dynamics. Fairhurst and Da Costa-Abreu [10] conducted a study using a multiclassifier system on the GREYC-keystroke database [12], and achieved an accuracy for gender prediction

[1] It records the response at the palm of the hand while sending a low voltage electrical current through the body from the other palm.

of 95%. Giot et al. [13] conducted a similar study using fixed-text input for gender prediction and reported an accuracy of 91%. They also reported that traditional keystroke authentication systems had an accuracy increase of 20% when combined with the user's gender prediction model. These studies achieve impressive accuracy for gender classification, but further research is required to determine if these results can be generalized to different sets of keyboard data that are not fixed, as well as to different types of keyboard interfaces. In addition, authentication systems based on keyboard dynamics may not be suited to new graphical password interfaces (see Biddle et al. for a survey of these interfaces [5]).

Mouse dynamics have been employed as a means of reauthentication to discriminate the identities of web browser users [21]. Ahmed et al. [3] used neural networks to learn a user's mouse dynamics in a specific environment while performing continuous identity authentication. Hamdy and Traore [14] combined mouse dynamics with cognitive measures of visual search capability and short term memory to create a static user verification system. These studies highlight the utility of using mouse biometrics in user re-authentication; however their findings are limited to identity authentication and have not been generalized to other purposes. To the best of our knowledge, no previous studies have reported the use of mouse biometrics to classify users' gender.

3 Methodology

This section describes the apparatus and method used for data collection. The data analysis procedures used to calculate and evaluate movement metrics are also described in this section.

3.1 Data Collection

There are 94 participants (45 men and 49 women) aged between 17 and 48 years participated in this study. The participants consist of students, faculty, and staff who were all experienced computer mouse users. The male and female participants did not differ statistically with respect to prior computer use experience or age.

All participants were seated in a static non-reclining chair in front of a computer monitor with the right hand resting comfortably on the same mouse and table surface used by all participants. Participants were instructed to find a seating location and arm posture in which moving the mouse would feel the most natural to them. They were requested to maintain this posture while conducting all experiment trials.

Raw mouse movement data were collected using an application implemented with the processing programing language. The same home (starting point) target was used on all trials and was displayed within an application window. Once a participant positioned the cursor on the home target and clicked the mouse button, this target was hidden and a new endpoint target was displayed. The screen position of the mouse was recorded at a rate of approximately 100 Hz with each data point consisting of a timestamp, the x screen coordinate, the y screen coordinate, and a tag that identified what type of a movement event was recorded. The movement events consisted of a standard movement event (mouse stationary or in motion without the left button being depressed), a target click event (left mouse button depressed while the mouse cursor is located inside the target area), a click event (left mouse button depressed while the cursor is outside of the target area), and a new target event (a new target displayed and the location and size of the target are recorded, instead of the mouse location).

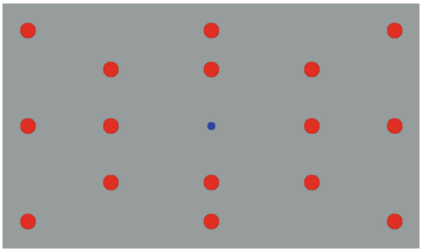

Fig. 3. Illustration of screen target positions for movements of mouse cursor. Home target located in center of window. All endpoint target positions are displayed in this diagram. (Color figure online)

The display window consisted of a rectangular frame (1680 px $\times$ 1050 px) displayed on a 45×30 cm computer monitor. As Figure 3 shows, the home target consisted of a blue 30 px radius circle located in the center of the display window. All endpoint targets were displayed as red circles and consisted of one of two possible target sizes (30 px or 60 px radius) located at one of 16 possible locations. The endpoint target locations varied in their direction of approach and in their distance from the starting target position.

Each participant was instructed to move the mouse cursor from the home target to the endpoint target. Once the participants had located the cursor in the home target circle, they were requested to click the mouse button to start the trial. The participants were instructed to only pick up the mouse when readjusting the starting position of their hands on the table, during which they were moving the screen cursor back to the home target. Each participant conducted a sequence of 32 practice trials that consisted of all 32 possible combinations of target size, target distance, and angle of approach as describe above. After successfully completing the practice trials, each participant then performed four blocks of 64 movement trials with each block of trials consisting of a random sequence of two trials for each combination of the 16 target locations and 2 target sizes. The participants were allowed to take a short rest after completing each block of movement trials.

3.2 Movement Metrics

The profiles of distance and velocity were extracted from the raw data of each movement trial. These profiles were used to calculate ten temporal metrics that distinguish aiming movements and button presses. The spatial trace of each movement was smoothed, and then six spatial metrics were calculated to highlight differences in the trajectory of each movement. Five accuracy metrics were also calculated for each mouse movement. Following the naturalistic approach, the choices of these metrics were guided by previous empirical research on gender differences in aiming movements that have used the same or similar metrics [4,6,9,15,23,24,27]. For example, researchers have reported that men are quicker at perceiving object location, faster in their movements, rely less on visual guidance of the ballistic omponent of the movement, perform less visual corrections towards the endpoint of the movement, and are less accurate when they reach the endpoint of the movement. Some additional metrics were calculated, because prior empirical research would imply gender differences are possible for these mouse metrics even if they were not reported in the actual studies. For example, males and females differ in their grip postures of the mouse and positioning of the finger over the mouse button [16,18,32], implying that gender differences could exist for metrics influenced by these grip postures.

3.2.1 Profiles

The distance profile was calculated from the Euclidean distance traveled between consecutive movement events, and smoothed using a Kolmogorov-Zurbenko (KZ) filter. The KZ filter belongs to the low pass filter class, and is a series of k iterations of a moving average filter with a window size of m, where m is a positive odd integer. In other words, the KZ filter repeatedly runs a moving average filter with the initial input being the original data and the result of the previous run of the moving averages as the subsequent inputs. With this in mind, the first iteration of a KZ filter over a process $X(t)$ can be defined as:

$$KZ_{m,k=1}\left[X(t)\right] = \sum_{s=-2(m-1)/2}^{2(m-1)/2} X(t+s)\frac{1}{m},$$

the second iteration as:

$$KZ_{m,k=2}\left[X(t)\right] = \sum_{s=-2(m-1)/2}^{2(m-1)/2} KZ_{m,k=1}\left[X(t+s)\right]\frac{1}{m},$$

and so on. In this study, we set m to 11 and k to 3, respectively. The value of $m = 11$ was chosen such that the window over which the data is averaged would correspond to 100 ms or more. Thus, the window can cover a period of time with an intentional movement since smaller ones are likely to be just jitters. The value 11 was chosen, instead of 10, because the value of m needs to be odd. The value $k = 3$ was chosen because 3 was the smallest value that produced a smooth curve.

The velocity profile was then calculated from sets of pairs (t, v_t), where v_t is the average velocity in pixels per millisecond (px/ms) over the time interval between t and the time at which the previous data point was recorded.

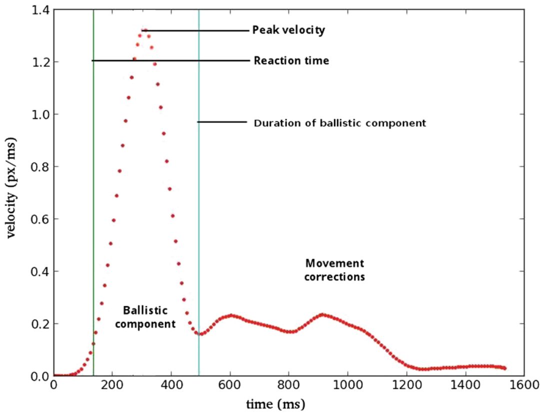

Fig. 4. Example of a velocity profile with various temporal metrics illustrated.

Aiming movements generally produce velocity profiles that are composed of one large peak (peak velocity) called the ballistic component that is followed by zero or more smaller peaks that reflect sub-movements used to position the cursor over the final target position (refer to Fig. 4). The velocity profile was used to calculate some of the 10 temporal features of the mouse dynamics recorded from each participant.

3.2.2 Temporal Movement and Button Press Metrics

- *Reaction time (RT):* the time interval from the moment the endpoint target appears on the screen until the participant initiates a movement towards it. The onset of the movement was determined to begin at the point when movement velocity exceeded 7% of the peak velocity for the ballistic component (refer to Fig. 4). Various methods were tested for determining the beginning point of movements, including measuring the slope of the velocity profile, pixels moved during consecutive time steps, and the percentage of peak velocity exceeded. All methods were tested using a visual inspection of a randomly selected group of trials and a set of known edge cases. Through this testing, we found that using the percentage of peak velocity exceeded with a value of 7% was the most effective solution.
- *Peak velocity (PV):* the maximum velocity value found for the ballistic component of the movement (refer to Fig. 4).
- *Time to peak velocity (TPV):* the time interval from the beginning of the movement until the peak velocity was reached (refer to Fig. 4).

– *Duration of ballistic component (DB):* the time interval from the beginning of the movement until the first local minima on the velocity profile following the peak velocity (refer to Fig. 4).
– *Shape of the velocity profile (SV):* a measure of the symmetry of the ballistic component, which is calculated by dividing the time to the peak velocity by the duration of the ballistic component (refer to Fig. 4).
– *Proportion of the ballistic component (PB):* the proportion of the movement time taken up by the ballistic component, which is calculated by dividing the ballistic component duration by the movement time (refer to Fig. 4).
– *Number of movement corrections (NC):* the total number of observed local maxima in the velocity profile after the ballistic component has been completed (refer to Fig. 4).
– *Time to click (TC):* the time interval between the arrival at the endpoint of the movement and the pressing of the mouse button.
– *Hold time (HT):* the amount of time the user held the mouse button down after the endpoint of the movement was reached.
– *Movement time (MT):* the time interval from the beginning of the movement until the endpoint of the movement.

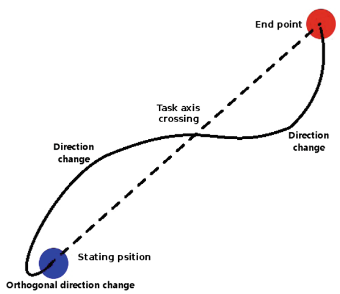

Fig. 5. Example of a mouse trajectory to illustrate differences between three movement change metrics with task axis drawn in a dashed line.

3.2.3 Spatial Movement Metrics

These metrics are calculated from the spatial trajectory traveled by the mouse cursor for reaching the endpoint of the movement.

– *Path length (PL):* the total distance traveled by the mouse cursor during the trial. It is calculated as follows:

$$\sum_{t=1}^{T} \Delta d_t$$

where T is the total number of the trial, and Δd_t represents the distance traveled between time t and time $t-1$.

- *Path length to best path ratio (PLR):* the value of the path length divided by the length of the shortest path between the start and endpoints of the movement.
- *Task axis crossings (TXC):* the number of times that the movement path crossed the task axis. The task axis is defined as a straight line between the home target and the endpoint target (refer to Fig. 5).
- *Movement direction changes (MDC):* the number of times the movement changed direction perpendicular to the task axis (refer to Fig. 5).
- *Orthogonal movement changes (OMC):* the number of times the movement changed direction parallel to the task axis (refer to Fig. 5).
- *Movement variability (MV):* the standard deviation of the distance of the movement path to the task axis. This metric measures the spatial consistency of the movement path.

3.2.4 Movement Accuracy Metrics

These metrics represent how closely a participant came to clicking the center of the endpoint target.

- *Absolute error (AE):* absolute error corresponds to the Euclidean distance between the endpoint of the movement and the center of the endpoint target.
- *Horizontal error (HE):* the difference in the horizontal (x) coordinates between the endpoint of the movement and the center of the endpoint target. Negative errors reflect undershooting the target location whereas positive errors reflect overshooting the target location.
- *Vertical error (VE):* the difference in the vertical (y) coordinates between the endpoint of the movement and the center of the end position target. Negative errors reflect undershooting the target location whereas positive errors reflect overshooting the target location.
- *Absolute horizontal error (AHE):* the absolute value of the difference in the horizontal coordinates between the endpoint of the movement and the center of the endpoint target.
- *Absolute vertical error (AVE):* the absolute value of the difference in the vertical coordinates between the endpoint of the movement and the center of the endpoint target.

These defined errors are illustrated in Fig. 6, where an absolute error consists of Euclidean distance between the end of a movement and the center of an endpoint target. The horizontal error corresponds to the difference in the x coordinates of the movement endpoint and the center of the endpoint target. The vertical error corresponds to the difference in the y coordinates of the movement endpoint and the center of the endpoint target. In both cases, a negative value depicts undershooting and a positive value depicts overshooting.

3.3 Data Filtering

Before calculating the movement metrics for each participant as described above, the movement data were filtered to remove invalid trials where mouse movements

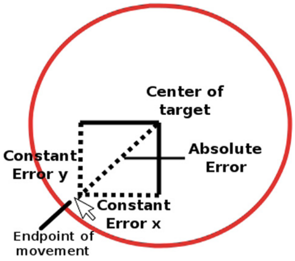

Fig. 6. Graphical depiction of movement accuracy metrics.

did not fall within the acceptable criteria for successful movement recording. The trials in which mouse movements clearly left the designated screen window were rejected, as well as the trials where the reaction times were less than 150 ms. This value of 150 ms was chosen, because the lower end of human reaction time is 100 ms. However, the method of determining the start of the movement is not perfect and causes some false positives. The same visual testing for determining the movement onset was used here, and we found that the value of 150 ms made a good balance between the false positive ratio and the false negative ratio while determining if the reaction time value was realistic. Only 4% of data points were rejected for these reasons across those more than 24,000 trials recorded.

4 Model Design

The gender classification model results from a two-step procedure of statistical analyses. The first step involves conducting least-squares multiple regressions to determine the effects of target parameters (target size, horizontal distance, and vertical distance) on movement metrics for each participant. The resulting unstandardized regression coefficients provide a movement signature for each participant, which will be used to distinguish the corresponding participant's gender. The second step involves conducting logistic regressions to select the statistical model that most accurately classifies participants by gender.

4.1 Mouse Signatures

Traditional analyses of mouse biometrics usually rely on a single aggregate indicator (e.g., average) for each metric. Unfortunately, previous studies have shown that this approach may be ineffective. For example, in the study conducted by Rohr [23], men were shown to have their accuracy reduced as a target was made smaller and placed further away, whereas women were more consistent with their accuracy. By simply taking the average accuracy, the gender difference would be diminished or lost since the lower values counteract the higher values. Thus, it is imperative to find a new way to produce features that capture not only the actual values observed in the data, but also the amount of changes caused by the target parameters. Our approach involves a more detailed analysis that incorporates the effects of target parameters on these mouse metrics. The effects of target parameters on the mouse metrics were quantified by unstandardized regression coefficients obtained from a multiple linear regression analysis with least squares fitting conducted for each metric. Multiple regression analyses predict the scores of a dependent variable y by fitting a straight line defined by a set of independent variables $\{x_1, x_2, x_3, ...\}$ to a set of known data points $(y_i, x_{1,i}, x_{2,i}, ...)$ such that it satisfies the equation:

$$y_i = a + b_1 x_{1,i} + b_2 x_{2,i} + ... + b_n x_{n,i} + \varepsilon_i,$$

where a and b_k are unknown constants that are estimated, and ε_i is the residual defined as the vertical deviation of the known data to the estimated line. If the estimated line is a perfect fit, all values of ε are zero.

The least squares fitting method estimates the values of a and b_k by reducing the squares of the residuals such that the following equation is minimized:

$$\sum_{i=1}^{r} \varepsilon_i^2 = \sum_{i=1}^{r} [y_i - (a + \beta_1 x_{1,i} + b_2 x_{2,i} + ... + b_n x_{n,i})]^2.$$

Three target parameters were chosen as predictor variables for these multiple regressions: the size of the endpoint target, the vertical distance between the home and endpoint targets, and the horizontal distance between the home and endpoint targets. The target distance was measured in separate horizontal and vertical components, because prior research suggests that these components should be the most influential on aiming movements rather than more complex combinations of the angle of approach and distance moved [29]. Absolute values were used for the distances traversed because previous research also suggests that the direction of movement (left vs. right and up vs. down) does not affect movement metrics as much as whether it is just a vertical movement or a horizontal movement [7,8]. Consequently, the size and sign of the regression coefficients for the distance variables simply represent how much of an effect, moving vertically or moving horizontally, had on the predictability of a metric.

For each metric recorded, three regression coefficients and the intercept value were provided to highlight the effect of these target parameters on the metric. For example, if the peak velocity (PV) was used as the dependent variable, four values were provided for this metric (intercept value PV_{const}, regression coefficient for horizontal distance moved PV_{horz}, regression coefficient for vertical distance moved PV_{vert}, and regression coefficient for target size PV_{size}). This results in a metric vector for the peak velocity that specifies the following equation:

$$PV = PV_{const} + PV_{size}(size) + PV_{vertD}(vertD) + PV_{horzD}(horzD).$$

It was expected that these regression variables would better reveal gender differences in the metrics. This assumption is supported by 4-way ANOVAs (gender $\times$ target size $\times$ distance $\times$ angle of approach) that were conducted for each metric. The significant results of these ANOVAs are summarized in Table 1. These results clearly show that many of the metrics revealed consistent target parameter effects, and these effects could be mediated by gender.

Table 1. Significant main effects and interactions found for 4-way ANOVAs (Gender × Distance × Angle of approach × Target size) conducted for each metric.

Variable	Significant effects
Reaction time	Gender, Distance, Size, Angle, Distance × Angle, Gender × Distance × Size × Angle
Movement time	Distance, Size, Angle
Hold time	Gender, Size, Angle
Time to peak V	Distance, Size, Angle, Distance × Angle, Gender × Size × Angle
Peak velocity	Distance, Size, Angle, Distance × Angle
T ballistic comp	Distance, Angle
Shape of velocity profile	Distance, Angle, Distance × Angle
Ballistic prop	Distance, Size, Angle, Gender × Size, Distance × Size, Distance × Angle, Size × Angle
N of corrections	Distance, Size, Angle, Distance × Size, Distance × Angle, Size × Angle
Time to press	Size, Angle
Path length	Distance, Size, Angle, Gender × Size, Distance × Angle, Gender × Size × Angle
Path L best ratio	Distance, Size, Angle, Size × Angle
Axis crossings	Distance, Angle, Distance × Angle
Direction changes	Distance, Size, Angle
Orthog changes	Distance, Size, Angle, Distance × Angle, Size × Angle
Movement var	Distance, Angle, Distance × Gender, Distance × Angle, Gender × Distance × Angle
Index of dIff	Distance, Size, Angle, Distance × Size, Distance × Angle, Size × Angle, Distance × Size × Angle
Index of performance	Distance, Size, Angle, Size × Angle
Horizontal error	Size, Angle, Size × Angle, Gender × Distance × Angle
Vertical error	Size, Angle, Size × Angle
Absolute error	Distance, Size, Angle, Size × Angle

4.2 Gender Prediction Model

The second step in developing a gender prediction model involves with the input of the metric variables obtained from each participant in a logistic regression to predict the gender of a participant. The logistic regression is often used for classification when dependent variables have binary values. The curve used in this type of regression is an S shaped curve asymptotically tapered between 0 and 1 and is derived from the following linear relation:

$$logit(P) = \alpha + \beta_1 x_1 + \beta_2 x_2 + ...,$$

where $logit(P)$ refers to the natural logarithm of the odds function defined as follows:

$$logit(p) = ln(odds) = ln\left(\frac{P}{1-P}\right).$$

This function can then be substituted into the original linear relation and be solved for P giving the formula:

$$P = \frac{e^{\alpha+\beta_1 x_1+\beta_2 x_2+\ldots}}{1+e^{\alpha+\beta_1 x_1+\beta_2 x_2+\ldots}},$$

where P is the probability that the dependent variable has the outcome coded as 1 given the values of x_i.

The values of constant α and coefficients β_i are determined by maximizing the conditional probability of the observed data, given the parameters used as predictors. An initial model is constructed with arbitrary values for the coefficients, and the conditional probability is evaluated. The coefficients are then modified in order to increase this probability, and the procedure is repeated until the model converges or a maximum number of iterations are reached. A maximum of 20 iterations were allowed to determine the values of the coefficients, and the results lead to a threshold value of 0.5 (i.e., whose values above 0.5 were considered as male and whose values no larger than 0.5 were considered as female).

5 Evaluation

The accuracy of the proposed approach for classifying a user' gender was evaluated on both labeled and unlabeled data. The labeled data consisted of the full data set, while the unlabeled data test was performed with 70% of the participants used as the training set and the remaining 30% of participants used as the test set (Table 2).

Table 2. Accuracy of predicted results. Labeled set refers to the full data set used in Sect. 4.1. Labeled 70% and unlabeled 30% refer to the training set and test set used in Sect. 4.2, respectively.

Set	Full set			Outliers removed		
	Labeled	Labeled 70%	Unlabeled 30%	Labeled	Labeled 70%	Unlabeled 30%
Male	91.1%	83.9%	57.1%	100%	100%	71.4%
Female	87.8 %	91.2%	86.7%	100%	100%	80.0%
Total	89.4%	87.7%	72.4%	100%	100%	75.9%

5.1 Labeled Data Analysis

In this section, we verify how well a model may be fit the data and the accuracy of such a model on users who have been sighted before. We also use this step to identify any users with unusual characteristics as outliers. The logistic regression model was tested on all 94 participants, but given the very large number of predictor variables (21 metrics $\times$ 4 metric features = 84 predictor variables) only smaller sub-sets of predictor variables were actually tested. The first subset of predictor variables was determined by testing each metric separately. The four features of each metric were tested as a single group separate from the features of the other metrics. The statistical significances ($p < 0.05$) of each metric's variables for predicting gender determined if these variables were included in the first sub-set of predictor variables. The significant predictors included in this subset were: $\{HT_{const}$, PV_{horz}, PB_{size}, TC_{const}, TC_{horz}, MDC_{const}, MDC_{horz}, MDC_{size}, $AE_{const}\}$. To improve the overall accuracy of this model, additional predictor variables were included while providing a moderate level of statistical significance ($p < 0.1$) in predicting gender when each metric was tested separately. Two additional variables were included to this sub-set of predictor variables: PB_{const} and PLR_{vert}. The amount of explained variance in gender classification using these two subsets of variables was 0.532 according to the Nagelkerke pseudo r-squared measure, and the classification accuracy based on this model was 75.5%.

The first subset of predictor variables was reduced from a total number of 84 to 9 by examining each metric's predictive power one metric at a time. However, a better subset of predictors may be possible if multiple metrics are included in the initial logistic regression model. One way to reduce the number of tested metrics is to only include those metrics that can characterize significant gender effects from the previously conducted 4-way ANOVAs. These findings highlight the metrics that show consistent gender differences or interactions of gender with target parameters. We also included those metrics published by other researchers with significant gender effects. The logistic regression model was tested again with a new subset of predictors that included the four variables for each of these metrics: $\{RT$, HT, TPV, PB, PL, MV, AE, HE, TC, PV, AHE, AVE, $VE\}$. The 52 predictor variables in this subset were added to the original subset with a stepwise method, and the following 10 new variables were revealed as significant predictors: $\{RT_{size}$, RT_{horz}, RT_{vert}, TPV_{vert}, MV_{const}, MV_{vert}, MV_{horz}, PV_{const}, PV_{vert}, $VE_{const}\}$. The amount of explained variance after the addition of these variables to the final model was 0.676, and the resulting classification accuracy was 89.4%.

We now test the effects that outliers had on the model. Five users were identified as having scores that were more than two standard deviations away from the mean. These are likely users with mouse movement characteristics that do not entirely fit the average for their gender, since there can be an overlap of physical characteristics between the two populations and such an overlap affects the features being used. After the removal of these outliers, our model can discriminate the gender of the remaining 89 participants with an accuracy of

100%. It is difficult to uncover the actual causes for these outliers, and they can occur for a variety of reasons including, but not limited to, distraction or injury. In a real application, one would likely test for outliers at input time, and if an outlier is detected, the user would be asked to re-do the input trials in the case of a one time authentication. However, identifying the best methods to handle outliers is beyond the scope of this paper.

5.2 Unlabeled Data Analysis

To evaluate the accuracy of our approach on unlabeled data, the movement data from 65 randomly selected participants were used as the training set to create the logistic regression model. And the model was then tested on the movement data from the remaining 29 participants who comprised the test set. The same variable selection procedure was followed with the unlabeled data as the one used for the labeled data, except that substantially fewer participants were involved in these selections.

The statistically significant predictors determined for subset one were: HT_{const}, TC_{horz}, MDC_{const}, MDC_{size}, MDC_{horz}, AE_{const}, AHE_{const}, AHE_{horz}, RT_{const}, PB_{size}, and VE_{vert}. Six of these predictor variables were consistent with the selections based on the full data set (labeled data). The fit of this model was tested on the training set and accounted for 0.449 of the explained variance in predicting gender with a correct classification of 76.9% of the participants in the set. The second subset included the following predictor variables: $\{PV_{const}, PV_{vert}, PV_{horz}, MV_{vert}, RT_{size}, RT_{vert}, RT_{horz}\}$. All seven variables were included in the subset of the predictors obtained previously with the full data set (labeled data). This overlap shows that this feature selection method produces a set of features close to what is expected based on research in other fields. On the other hand, what can be observed over the entire set may still have sensitivity to the training set, which one should be careful of when fitting the model. The fitness of this model with the combined subsets was tested on the training set and accounted for 0.579 of the explained variance in predicting gender. This final model was tested on the test set and was able to achieve a gender classification accuracy of 72.4%. After removing the outliers identified previously in the labeled data analysis, the test set was then classified with a 75.9% accuracy. These results suggest that outliers have a visible effect on the classifier, but the negative impact is relatively small.

6 Discussion

Men and women differ naturally, both physically and psychologically. The development of computer security tools can take advantage of these natural differences by focusing authentication procedures on these differences. This study used the naturalistic approach to successfully classify male and female participants by measuring the temporal, spatial, and accuracy characteristics of their mouse movements while evaluating how these mouse metrics were affected by target parameters.

The measurement of one such metric, movement accuracy, will be used to exemplify this approach to the biometric analysis of mouse dynamics. Previous research with aiming movements has revealed gender differences in the spatial accuracy of these movements with women being on average more accurate than men [4,23]. However, this gender difference is actually more complicated than one suggested by simply comparing average errors, because target parameters (target size, distance moved, and direction of movement) can also differentially affect the movement accuracy of men and women [23]. In support of this premise, our study also found complex interaction effects of gender and target parameters on spatial error. Consequently, rather than just recording the mean accuracy of each participant's movements, a multiple regression analysis was conducted to predict spatial error using target parameters (size, horizontal distance, vertical distance) as predictor variables.

This novel approach to biometric analysis comes with some cost, because there are now four variables representing each metric's potential contribution to the prediction model. Given the relatively large number of movement features already required by our approach, a large number of predictor variables could be introduced to discriminate the gender of a participant using logistic regressions. Therefore, two criteria were followed to reduce the set of predictor variables for testing: (1) each metric was tested individually and only those variables that were significant predictors of gender in these tests were included in the first subset of predictors, (2) all the metrics that produced significant ANOVA gender effects and those with gender effects suggested in prior research were included in a second subset. Our logistic regressions produced correct classification of a participant's gender at a rate of 89.4–100% for the labeled data and 72.4–75.9% for the unlabeled data. These results are very promising given the limited range of values provided for each target parameter in this study.

The optimal classification accuracy was achieved after removing outliers from the labeled data set and from the training data set for the analysis of unlabeled data. It is unclear why a few participants had such discrepant mouse metrics, and further research is needed to rule out the possibility of introducing user behavioral outliers into data collection and evaluation. However, their effects on the unlabeled data were minor, indicating that they do not have a large impact on classifying previously unseen users.

Once the recording accuracies of the movement metrics have been established, the current procedure has very low computational overheads because it relies on simple statistical models for computing predictor variables and gender classification. A client machine can collect the raw movement data and then send it to a server for feature extraction and prediction of gender with minimal overhead, and relatively low latency for the client. Consequently, static and continuous authentications are viable options with this approach. In fact, real-life mouse movements that are not constrained to an experimental manipulation, as was the case in the current study, should provide a larger range of target parameters and therefore better predictive accuracy. A larger, more diverse data set of participants would also facilitate the testing of this approach, because the majority of the participants in the current study were highly educated undergraduate college students.

7 Future Work

A direct application of this method to be explored is the generalization of this method across computer platforms with different hardware. One major advantage of the naturalistic approach to biometric analysis is that predictive models based on natural differences are assumed to have a universal, biological basis, and therefore, should be more generalizable than traditional data driven approaches to biometrics analysis. Accordingly, the gender classification model formulated in the current paper should generalize to other populations of computer-users (e.g., other countries, different education backgrounds), and also, be somewhat independent of the computer-user environments where the mouse data are collected (e.g., table height, table surface, type of mouse etc.). A comparison of the classification success found in the current study for labeled and unlabeled data provides some support for this assumption. When participants were classified using a model based on another group's data (unlabeled data) there was still a reasonable rate of classification success (72.4–75.9%) albeit with some drop in performance from a completely labeled set of data (89.4–100%). Future research could examine this generalization prediction using different computer work stations and cross-cultural tests of classification success.

8 Conclusion

This paper proposes a naturalistic approach for gender classification of computer users based solely on their mouse movements. The design rationale of our approach lies in the observation that men and women differ naturally in how they make mouse movements. We defined a series of temporal, spatial, and accuracy metrics to quantify the mouse movement differences between male and female users. In particular, we identified the metrics related to peak velocity, length of the deceleration phase, target accuracy, finger posture, and reaction time are relevant to gender classification. There were 94 volunteers participated in this study, and a mouse signature was created for each participant. We evaluated the efficacy of our approach for gender classification by conducting binary logistic regression tests, and achieved promising results.

References

1. Anthropometric Reference Data for Children and Adults: United States, 2007–2010. Number 252 in 11. U.S. Department of Health (2012)
2. Ahmed, A., Traore, I.: Anomaly intrusion detection based on biometrics. In: Information Assurance Workshop, IAW 2005. Proceedings from the Sixth Annual IEEE SMC, pp. 452–453, June 2005
3. Ahmed, A., Traore, I.: A new biometric technology based on mouse dynamics. IEEE Trans. Dependable Secure Comput. **4**(3), 165–179 (2007)
4. Barral, J., Bettina, D.: Aiming in adults: sex and laterality effects. Laterality **9**(3), 299–312 (2004)
5. Biddle, R., Chiasson, S., Van Oorschot, P.: Graphical passwords: learning from the first twelve years. ACM Comput. Surv. **44**(4), 19:1–19:41 (2012)
6. Buchman, A.S., Leurgans, S., Gottlieb, G.L., Chen, C.H., Almeida, G.L., Corcos, D.M.: Effect of age and gender in the control of elbow flexion movements. J. Mot. Behav. **32**(4), 391–399 (2000)
7. Dounskaia, N.V., Ketcham, C.J., Stelmach, G.E.: Influence of biomechanical constraints on horizontal arm movements. Mot. Control **6**(4), 366–387 (2002)
8. Dounskaia, N.V., Swinnen, S.P., Walter, C.B., Spaepen, A.J., Verschueren, S.M.: Hierarchical control of different elbow-wrist coordination patterns. Exp. Brain Res. **121**(3), 239–254 (1998)
9. Jiménez-Jiménez, F.J., et al.: Influence of age, gender in motor performance in healthy subjects. J. Neurol. Sci. **302**(12), 72–80 (2011)
10. Fairhurst, M., Da Costa-Abreu, M.: Using keystroke dynamics for gender identification in social network environment. In: 4th International Conference on Imaging for Crime Detection and Prevention 2011 (ICDP 2011), pp. 1–6, November 2011
11. Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. J. Exp. PSychol. **74**, 381–391 (1954)
12. Giot, R., El-Abed, M., Rosenberger, C.: Greyc keystroke: a benchmark for keystroke dynamics biometric systems. In: Proceedings of the 3rd IEEE International Conference on Biometrics: Theory, Applications and Systems, BTAS 2009, Piscataway, NJ, USA, pp. 419–424. IEEE Press (2009)
13. Giot, R., Rosenberger, C.: A new soft biometric approach for keystroke dynamics based on gender recognition. Int. J. Inf. Technol. Manage. **11**(1/2), 35–49 (2012). Dec
14. Hamdy, O., Traoré, I.: Homogeneous physio-behavioral visual and mouse-based biometric. ACM Trans. Comput. Hum. Interact. **18**(3), 12:1–12:30 (2011)
15. Hansen, S., Elliott, D.: Three-dimensional manual responses to unexpected target perturbations during rapid aiming. J. Mot. Behav. **41**(1), 16–29 (2009)
16. Johnson, P.W., Hagberg, M., Hjelm, E.W., Rempel, D.: Measuring and characterizing force exposures during computer mouse use. Scand. J. Work Environ. Health **26**(5), 398–405 (2000)
17. Karlqvist, L.K., Bernmark, E., Ekenvall, L., Hagberg, M., Isaksson, A., Rost, T.: Computer mouse position as a determinant of posture, muscular load and perceived exertion. Scand. J. Work Environ. Health **24**, 62–73 (2012)
18. Lee, D.L., McLoone, H., Dennerlein, J.T.: Observed finger behaviour during computer mouse use. Appl. Ergon. **39**(1), 107–113 (2008)
19. Monrose, F., Reiter, M.K., Wetzel, S.: Password hardening based on keystroke dynamics. In: Proceedings of the 6th ACM Conference on Computer and Communications Security, CCS 1999, New York, NY, USA, pp. 73–82. ACM (1999)

20. Phillips, J.G., Triggs, T.J.: Characteristics of cursor trajectories controlled by the computer mouse. Ergonomics **44**(5), 527–536 (2001)
21. Pusara, M., Brodley, C.E.: User re-authentication via mouse movements. In: Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security, VizSEC/DMSEC 2004, New York, NY, USA, pp. 1–8. ACM (2004)
22. Rasmussen, K.B., Roeschlin, M., Martinovic, I., Tsudik, G.: Authentication using pulse-response biometrics. In: The Network and Distributed System Security Symposium (NDSS), vol. 2 (2014)
23. Rohr, L.E.: Gender-specific movement strategies using a computer-pointing task. J. Mot. Behav. **38**(6), 431–437 (2006)
24. Rohr, L.E.: Upper and lower limb reciprocal tapping: evidence for gender biases. J. Mot. Behav. **38**(1), 15–17 (2006)
25. Soukoreff, R.W., MacKenzie, I.S.: Towards a standard for pointing device evaluation, perspectives on 27 years of fitts' law research in HCI. Int. J. Hum. Comput. Stud. **61**(6), 751–789 (2004)
26. Stefan, D., Shu, X., Yao, D.D.: Robustness of keystroke-dynamics based biometrics against synthetic forgeries. Comput. Secur. **31**(1), 109–121 (2012)
27. Teeken, J.C., Adam, J.J., Paas, F.G., van Boxtel, M.P., Houx, P.J., Jolles, J.: Effects of age and gender on discrete and reciprocal aiming movements. Psychol. Aging **11**(2), 195–198 (1996)
28. Thibbotuwawa, N., Hoffmann, E.R., Goonetilleke, R.S.: Open-loop and feedback-controlled mouse cursor movements in linear paths. Ergon. Int. J. Res. Pract. Hum. Factors Ergon. **55**, 476–488 (2012)
29. Thompson, S.G., McConnell, D.S., Slocum, J.S., Bohan, M.: Kinematic analysis of multiple constraints on a pointing task. Hum. Mov. Sci. **26**(1), 11–26 (2007)
30. Wahlstrm, J., Svensson, J., Hagberg, M., Johnson, P.W.: Differences between work methods and gender in computer mouse use. Scand. J. Work Environ. Health **26**(5), 390–397 (2000)
31. Yampolskiy, R.V., Govindaraju, V.: Behavioural biometrics: a survey and classification. Int. J. Biometrics **1**(1), 81–113 (2008)
32. Yang, J.-F., Cho, C.-Y.: Comparison of posture and muscle control pattern between male and female computer users with musculoskeletal symptoms. Appl. Ergon. **43**, 785–791 (2011)