User Evaluation of Lightweight User Authentication with a Single Tri-Axis Accelerometer

Jiayang Liu Lin Zhong
Dept. of Electrical & Computer Engineering
Rice University, Houston, TX, USA
{jiayang,lzhong}@rice.edu

Jehan Wickramasuriya Venu Vasudevan
Software Platforms Research
Motorola Inc.,Schaumburg, IL, USA
{jehan, venu.vasudevan}@motorola.com

ABSTRACT
We report a series of user studies that evaluate the feasibility and usability of light-weight user authentication with a single tri-axis accelerometer. We base our investigation on uWave, a state-of-the-art recognition system for user-created free-space manipulation, or *gestures*. Our user studies address two types of user authentication: non-critical authentication (or identification) for a user to retrieve privacy-insensitive data; and critical authentication for protecting privacy-sensitive data. For non-critical authentication, our evaluation shows that uWave achieves high recognition accuracy (98%) and its usability is comparable with text ID-based authentication. Our results also highlight the importance of constraints for users to select their gestures. For critical authentication, the evaluation shows uWave achieves state-of-the-art resilience to attacks with 3% false positives and 3% false negatives, or 3% equal error rate. We also show that the equal error rate increases to 10% if the attackers see the users performing their gestures. This shows the limitation of gesture-based authentication and highlights the need for visual concealment.

Categories and Subject Descriptors
H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – Evaluation/methodology, Input devices and strategies, Interaction styles.

General Terms
Performance, Design, Human Factors, Security

Keywords
User study, authentication, gesture, accelerometer

1. INTRODUCTION
An increasing number of consumer electronics and mobile phones are equipped with accelerometers, enabling a device to "sense" how it is physically manipulated by the user. In this work, we use "gesture" to refer to such physical manipulation, including not only hand gesture as we commonly know but also any physical manipulation of the device like shaking and tapping.

Many have studied the use of accelerometers to recognize gestures [1-4]. Some recognizers, in particular our prior work uWave [1], allow the user to train the recognizer with as few as one single sample. Such recognizers provide an interesting opportunity for gesture-based user authentication, which is light-weight in terms of computing, form factor, and user engagement. For example, a user can "shake" a phone in a special way to log in or a TV remote to load personalized data. While many paradigms exist for user authentication, including password [5], biometrics [6-8], speech [9], and handwriting [10], accelerometer-based gesture recognition has its unique value for user authentication because of its low cost, high efficiency, and no form factor change. These properties make it highly suitable for implementation on resource-constrained devices, e.g. mobile phones and TV remotes.

MobileHCI'09, September 15 - 18, 2009, Bonn, Germany.
ACM 978-1-60558-281-8.

The goal of this work is to investigate the feasibility and usability of such gesture-based authentication using uWave, a state-of-the-art gesture recognition system based on a single tri-axis accelerometer [1]. We distinguish two different objectives of user authentication. For privacy-insensitive data, the objective of user authentication is to retrieve user-specific data instead of protecting them, e.g. personal profiles or personalized configurations on a TV remote shared by family members. In this case, accuracy and usability are dominating concerns. We call such user authentication *non-critical* and call the gestures *ID gestures*. On the other hand, there are also privacy-sensitive data, e.g. personal contacts stored in a mobile phone. In this case, the objective of user authentication is to protect the data against possible unauthorized access, or *attacks*. Therefore, resilience to attacks and usability are dominating concerns. In contrast to non-critical authentication, we call such authentication *critical* and the gestures *password gestures*.

In particular, we seek answers for the following questions.

- How accurate can non-critical authentication be?
- How difficult do users perceive memorizing and performing gestures, particularly in comparison to widely used text-based authentication?
- Because uWave allows users to create personalized gestures, how do users select their gestures for authentication? How can the authentication be improved with constraints in gesture selection?
- What tradeoffs between security and usability can accelerometer-based gesture recognition make for critical authentication?
- Since performing gestures is likely to be more visible than typing textual passwords, what is the impact of visual disclosure of password gestures?

To answer these questions, we design and conduct a series of user studies with controlled variations to examine accuracy, usability, and resilience against attacks in a comprehensive manner. The user studies involve 25 participants over one month.

For non-critical authentication, our user study demonstrates that uWave achieves average 98% accuracy with simple gesture selection constraints; a follow-up survey shows that the usability of uWave for non-critical authentication is comparable to the use of textual ID-based authentication. For critical authentication, we find 3% equal rate of false negatives (rejecting authentic users' gestures) and false positives (accepting attackers' gestures), or

equal error rate, can be achieved without visual disclosure, meaning the attacker does not see the owner's password gesture performance. Note that equal error rate is a standard performance measure of a classifier. The lower equal error rate, the more accurate is the classifier. We also find showing the owner's performance to the attacker, or visual disclosure, increases the equal error rate to 10%. Therefore gesture-based authentication can be used only when strict security is either unnecessary or can be achieved through combination of gesture-based authentication and traditional methods. Our evaluation highlights the need to conceal the gesture performance. Our analysis also shows the potential to achieve a lower equal error rate through adaptation to users.

Our work is the first publicly reported study that extensively evaluates the usability and feasibility of accelerometer-based authentication with user-defined gestures. Recent work [11, 12] has studied *predefined* gestures as behavioral biometrics , i.e. who the user is, however, our work studies user-created gestures as behavioral *secrets*, i.e. what the user knows. Moreover, existing work requires a large number of training data and does not address usability.

The rest of the paper is organized as follows. We discuss related work in Section 2 and then present an overview of uWave and its Wii remote-based prototype in Section 3. We present two series of extensive user studies for non-critical and critical authentications in Sections 4 and 5, respectively. We discuss how to improve critical authentication and present our observation on the choices of gestures in Section 6 and conclude in Section 7.

2. RELATED WORK

Most user authentication methods are based on either what properties the user has, e.g. fingerprint [6], face [7] and iris [8], or what the user knows, e.g. password [5], or both, e.g. speaker verification [9] and handwritten signature recognition [10]. All these methods, however, require form factor modification or considerable computation and user engagement, unsuitable for operating small resource-constrained devices in a mobile manner. In contrast, accelerometer-based authentication allows free-space hand movement and does not require any form factor change to the device.

The work in [11-13] considers gesture as a behavioral biometrics that the user has and attempts to verify or recognize the user identity based on a fixed gesture performed by all participants, e.g. a simple arm swing in [11]. In contrast, we allow the user to create any physical manipulation of the device as the authenticating gesture. In other words, our authentication approach is based on both what the user "knows" and what properties the user has. As a result, our work investigates the human factors in gesture selection and the usability of customized gestures. The goal of [11] is similar to that of our critical authentication: to verify a claimed user identity. The authors showed about 4% equal error rate over long time but through adaptation with a large number of training samples [13], compared to 3% in our solution of critical authentication with a single training sample. Notably, the basic method in [13] has over 14% equal error rate when not as many training samples are used. Moreover, the authors did not investigate how robust their methods are against attackers imitating the user, which is an important issue our work investigates. The goal of [12] is similar to that of our non-critical authentication: to recognize a user out of a small number of users sharing a device. The work achieved an accuracy of about 95% with a large number of training samples, ten versus a single one with our method, and the user must perform a given gesture in a highly constrained manner, e.g. exact timing with real-time visual feedback. These are challenging requirements for implementation on resource-constrained smart objects in mobile computing. More importantly, the gestures performed by a participant were collected from the same day, while both [13] and our prior work [1] showed there were significant variations in how a user performs the same gesture over time. As a result, the result reported in [12] is likely to be overly optimistic. In contrast, our solution achieves 98% over a period of four weeks. The fundamental reason is that our solution allows the users to create their personalized gestures and therefore allows more distinct features in their gestures.

Related to our use of accelerometers, the work in [14] employed accelerometers to recognize the user with the gait pattern as a behavioral biometrics. Accelerometers have also been used to solve another security-related problem, pairing of two devices [15-19]. The approach is to produce a time series of acceleration as the shared secret between two devices. Such work, however, is very different from ours in their goal and scope.

3. GESTURE RECOGNITION BASED ON ACCELEROMETER

We next describe uWave, the gesture recognition system which our user studies are based on.

3.1 UWave: Personalized Gesture Recognizer

UWave bases the recognition on matching two time series of forces, measured by a single three-axis accelerometer. It employs a *template library* that stores one or more time series of known vocabulary gestures, often input by the user. The input to uWave is a time series of acceleration provided by a tri-axis accelerometer. The tri-axis accelerometer measures the acceleration it experiences on three orthogonal directions. Each time sample of the accelerometer reading is a trio of the acceleration along the three axes. uWave first quantizes acceleration data with multiple discrete levels to compress the data and filter out small noises. The same quantization applies to the templates too. uWave then employs dynamic time warping (DTW) to match the input time series against the templates of the gesture vocabulary and calculate the different between the two as a matching distance. It recognizes the gesture as the template that provides the smallest matching distance. The recognition results, confirmed by the user as correct or incorrect, can be used to adapt the existing templates to accommodate gesture variations over time. More information regarding uWave can be found in [1].

UWave plays different roles in non critical authentication and critical authentication. In non critical authentication, uWave identifies the best matching template from multiple templates created by different users and recognize the user as the identity behind the best-matching template. In critical authentication, uWave functions as a classic binary classifier. It calculates the matching distance between the input gesture and the template gesture representing the claimed user identity. If the matching distance is lower than a certain threshold, the input gesture is accepted as the claimed identity and otherwise rejected. For critical authentication, uWave may make two types of errors: false positive when it accepts an attacker's input; and false negative when it rejects the owner's input. A higher threshold leads to more false negatives (rejecting authentic user's gestures) so that less usable but less false positives (accepting attackers' gestures) so that more secure. By varying the threshold, one can obtain the receiver operating characteristic (ROC) curve, which quantitatively represents the

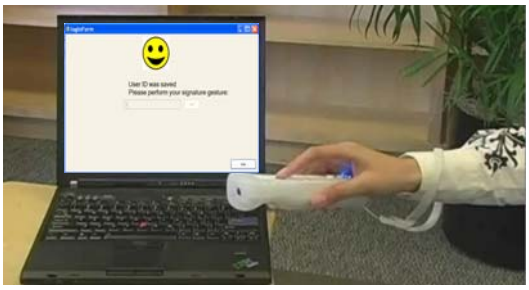


Figure 1: Wii remote based prototype of uWave: the Wii remote sends the acceleration data through Bluetooth to the laptop that runs the recognition algorithm

classifier's tradeoff between false positives and false negatives. An ideal classifier should be able to achieve both low false positive and low false positive rates.

In our implementation, the threshold for critical authentication is a portion of the *base distance*, calculated as the matching distance between the pre-recorded password gesture template and a still state acceleration sequence. The still state acceleration sequence is the acceleration data generated by the accelerometer when it is at rest on the Earth's surface for the same time duration as the gesture template.

3.2 Prototype for User Studies

We implement a prototype of uWave using Wii remote, as illustrated in Figure 1. A Wii remote has a built-in three-axis accelerometer from Analog Devices. It can send the acceleration data and button actions through Bluetooth to a PC in real time. We implement uWave and its variations on a Windows PC using Visual C#. The implementation is about 300 lines of code. The prototype detects the start of a gesture when the 'A' button on the Wii remote is pressed; and detects the end when the button is released. While our prototype is based on the Wii remote hardware, uWave can be implemented with any device having a three-axis accelerometer of proper sensitivity and range as are those found in most consumer electronics and mobile devices.

3.3 Complexity and Accuracy

UWave gives out recognition result without perceptible delay in our experiments based on multiple platforms. We measured the speed of uWave implemented in C on multiple platforms. On a Lenovo T60 with 1.6GHz Core 2 Duo, it takes less than 2ms for a template library of eight gestures. On a T-Mobile MDA Pocket PC with Windows Mobile 5.0 and 195MHz TI OMAP processor, it takes about 4ms for the same vocabulary. Such latencies are too short to be perceptible to human users. We also test uWave on an in-house built sensor platform with an extremely simple 16-bit microcontroller, TI MSP430. The delay is about 300ms. While this may be perceptible to the user, it is still much shorter than the time a gesture usually takes so that should not impair user experience because the DTW matching can start as soon as the first reading of acceleration comes in.

Using the Wii remote based prototype, we evaluate uWave with a set of eight simple gestures identified by a VTT study [4] and with a library of 4480 gestures collected from eight participants over multiple weeks. The study shows that uWave achieves accuracy of 98.6% with simple template adaptation for *user-dependent* gesture recognition, as reported in [1]. Such recognition accuracy is competitive with the results reported in [4] as 98.9% using HMM with 12 training samples.

4. NON-CRITICAL AUTHENTICATION

User identity can provide convenience to retrieve non-sensitive user-specific data. For example, many of the interfaces of advanced TV remotes can be personalized for users' convenience. When a user takes control of the remote, he/she would like to retrieve his/her profile in a seamless manner. For such non-critical authentication, uWave identifies the best matching template from multiple templates created by different users and recognize the user as the identity behind the best-matching template.

4.1 Objectives

We aim to answer the following research questions for non-critical authentication.

- What accuracy can uWave system achieve in recognizing users based on user-created ID gestures?
- How usable is it? In particular, how challenging is it to memorize and perform an ID gesture, in comparison to conventional text ID-based authentication?
- Since users are allowed to create their own gestures, what constraints in ID gesture selection can be employed to improve the accuracy and usability?

4.2 Procedure

4.2.1 Participants and Training

We recruit 25 participants. They are undergraduate and graduate students from multiple universities in the USA, aged 18 to 32, 18 males and 7 females. They major in various disciplines, including Chemistry, History, Electrical Engineering, Computer Science, Mechanical Engineering, and MBA. Some have international background.

We break the participants into five 5-person groups, called A to E in the follows. We conduct the user study for each group using a similar procedure with controlled variations. Table 1 summarizes procedural difference for all five groups. Before the user study, participants are given instructions on how to use the Wii remote-based uWave prototype and are also provided with basic information regarding its template-matching mechanism. They all play with the prototype to get acquainted.

4.2.2 Selecting ID Gestures

Table 1 summarizes the different procedures and constraints in ID gesture selection for each group.

All participants in Group A attend the first session at the same time and are asked to agree on a set of gestures as their IDs for authentication. We suggest them not to choose simple gestures as identified by a VTT study for home appliance remote control [4] as shown in Figure 2, called VTT gestures in the rest of the paper; we further suggest the gestures to be shorter than five seconds. The participants are allowed to test the system to see whether it could recognize their selections or not: each of them input his/her gesture for a few times and the system gives them the recognition result immediately each time. This is designed to allow the participants to evaluate their collective choice of gestures. In case two gestures are highly confusing, they might have a chance to try new ones. Nevertheless, the first choices by all five participants are recognized with 100% accuracy in the session. As a result, they converge on the selection with only one attempt.

Table 1: Procedure variations with Group A to E in the user study for non-critical authentication

Group	Gesture Selection Constraint	Session Frequency	Duration (week)	Textual ID
A	**Collectively**	**Every day**	**1**	**No**
B	**Individually; No constraint; no rejection procedure**	**Every day in the first week, Every two days in the second week; Every three days in the last two weeks**	**4**	**Yes**
C	**Individually; have constraint; no rejection procedure**	**Same as Group B**	**4**	**Yes**
D	**Individually; have constraint and rejection procedure**	**Same as Group B**	**4**	**Yes**
E	**Individually; have constraint and rejection procedure**	**Every day**	**1**	**No**

Participants in Groups B to E select their gestures one after another, without knowing the choices of others in the same group. Such a scenario is common for shared devices with gradual user adoption and provides an important alternative to the collective selection in Group A.

In order to compare the usability of ID gestures to that of commonly used textual IDs, we also ask participants in Groups B to D to choose a textual ID from a given list at the beginning of the study. Each textual ID on the list is comprised of a common used word of 3 to 8 letters and a randomly generated digit from 0 to 9. It is important to note that the purpose of these textual IDs is to provide a consistent base for usability comparison, instead of to emulate the textual IDs used in real life.

For Groups B to E, we explore the impact of gesture selection constraints. The participants in Group B are free to choose any gesture as ID gesture. Four of them select very simple ones, similar to the VTT gestures. Those in Groups C to E are suggested to choose gestures more complex than the VTT gestures as shown in Figure 2. In addition, participants in Groups D and E have to compose gestures that can pass a rejection procedure: the first participant in a group is allowed to pick up any ID gesture; the rejection procedure will reject any subsequent ID gesture choices if the matching distance between them and existing ID gestures is below 50% of the average distance between each pair of the template gestures from Group A.

One input of the selected ID gesture by each participant is saved as his/her template. We ask the participants to note down their gestures on paper so that we can have the most accurate representation of the gestures. Although we may video tape the gesture performance, it is still difficult to accurately infer the path of the hand movement in free space. Noting down their gesture selections may help the participants remember the gestures. As observed from their gesture selections, however, many of the selected gestures are drawing of symbols or letters familiar to the participants and may have already been frequently used in their daily life. In this case, noting down the gesture should not significantly impact the results of the user studies. Figure 3 shows their choices. It is important to note that the gestures are generated by six-degree free-space movement and Figure 3 only provides a planar representation.

4.2.3 Collecting Gestures for Evaluation

The gesture collection spans over one week for Group A and E. On each day, we invite the participants back to the lab to verify

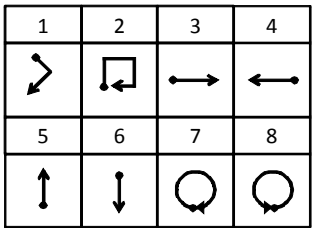


Figure 2: VTT gestures

themselves with their ID gestures. Each participant performs his/her ID gesture ten times and gets the recognition result immediately after each input, similar to what would happen with user recognition in reality. With the immediate feedback, the participant can adjust the next input in case of a recognition error.

The gesture collection for Groups B, C and D takes four weeks each. We invite participants back periodically with decreasing visit frequency to study the challenge of memorizing ID gestures. Participants visit us every day in the first week, every two days in the second week, and every three days in the last two weeks. In each visit, the participants perform their ID gesture 10 times and type in their textual IDs to login a laptop. They get the authentication result (success/fail) immediately after each input. If the average accuracy of a participant's inputs drops below 50% in one session, we replace his/her template using the latest input. Such template replacement is motivated by findings from our early work [1] that it helps uWave cope with gesture variations over time.

4.2.4 Surveys

We conduct a structured survey with participants at the end of the study to evaluate their subjective opinions on the usability of gesture-based authentication, which engenders two unique tasks: 1) memorizing the ID gesture and 2) performing it.

Our hypothesis to compare the difficulties of memorizing a gesture and a textual ID is:

> *H1: memorizing an ID gesture is as difficult as or more difficult than memorizing the pre-composed textual ID.*

Our hypothesis to compare the physical difficulty of performing a gesture and typing in a textual ID is:

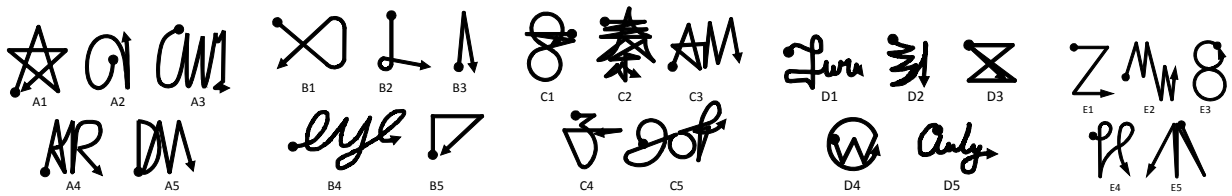


Figure 3: ID gestures for non-critical authentication by all 25 participants

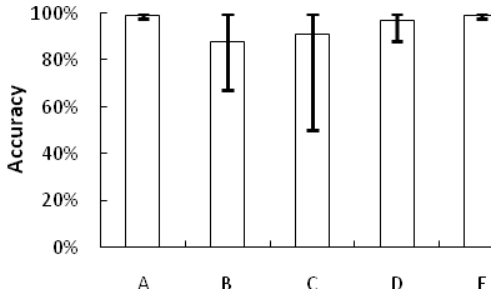


Figure 4: Recognition accuracy for non-critical authentication: range and average for Groups A to E

> *H2: performing an ID gesture is as difficult as or less difficult than typing in a pre-composed textual ID.*

In the survey, the participants are asked to rate 1) the difficulty of memorizing the gesture and the user ID in a 0 to 10 scale; 2) the difficulty of performing the gesture and typing in the textual ID in a 0 to 10 scale; 3) their agreement with two statements: "Memorizing a gesture is no more challenging than memorizing a textual ID" and "Performing a gesture is no more physically challenging than typing in a textual ID". There are also open-ended questions asking them to explain why.

4.3 Authentication Results

Figure 4 shows the average recognition accuracy in the first week for the five groups, since the data collection procedures in the first week are the same for all five groups. UWave achieves an average accuracy of 99.2% for Group A in which gesture complexity constraint is suggested and participants collectively select their ID gestures. Participants in Groups B to E select their ID gestures without knowing their peers'. The difference in the constraints of their gesture selection is detailed in Table 1. From Groups B to E, we observe the average accuracy increased from 88% to 99% due to gesture complexity constraints and the rejection procedure, which will be explained below.

4.3.1 Selection Constraints Improve Accuracy

The groups with complexity constraint and rejection procedure outperform the others: Group C has higher accuracy than Group B because of the complexity constraint; Groups D and E beat Group C because of the rejection procedure. The complexity constraint eliminates simple gestures which can be easily confused with each other; the rejection procedure guarantees enough difference between gesture templates. Both of these two conditions are important to help participants create usable ID gestures.

4.3.2 Template Replacement Improves Accuracy

In our prior work [1], we found that the same gestures performed by the same participant had significant variations over time. Thus template replacement is important to adapt to such variations. From the non-critical authentication study, we also observe that template replacement when the accuracy drops below certain threshold (50% in our study) can considerably improve accuracy. It enables users to overcome poorly inputted templates and adapt to performance variations over long time. In our study, B1 and B5 have low accuracy at the first three to five sessions, get their templates replaced when the accuracy drop below 50%, and achieve 92% and 98% accuracy respectively afterward. B3 has 50% accuracy at the very beginning but maintain 100% for the next a few sessions after template replacement. C2 and C5 have low accuracy (20~60%) in the first week, get templates replaced in the second week, and achieve 100% afterwards.

4.4 Usability Evaluation

We next evaluate the usability of gesture-based non-critical authentication in terms of difficulties in memorizing and performing an ID gesture. Figure 5 shows the group-wise average difficulty ratings. We analyze the ratings through hypothesis testing. We are aware of the debate on whether data from Likert scales should be viewed as interval-level data or ordered-categorical data [20]. In our survey, however, a visual analog scale with equal spacing between responses is shown to the participants and more than five levels are provided. Therefore, we believe it is propitiate to use parametric statistical test for analysis.

4.4.1 How difficult is memorizing gestures?

We use *independent two-sample t-test* to analyze the survey results. With data from all 25 participants, there is significant difference in the means of the difficulty rating for memorizing gesture and textual ID ($P = 0.04$). Since the P-value is smaller than our significance level (5%), we reject H1 as stated in Section 4.2.4, accept its alternative hypothesis, and conclude that memorizing a gesture is less difficult than a pre-composed textual ID.

4.4.2 How difficult is performing gestures?

We analyze the difficulty of performing gestures in a similar way with independent two-sample t-test. The result ($P = 0.24$) shows there is not enough evidence to reject H2 as stated in Section 4.2.4. In other words, our study does not find enough evidence to support that performing a gesture is more difficult that typing in a textual ID.

We hypothesize that participants' difficulty rating is negatively correlated to the recognition accuracy. For example, the participants in Group C consider performing a gesture much more difficult than typing in a textual ID. At the same time, Group C exhibits the largest variance in recognition accuracy as well as low average accuracy. To understand the correlation between accuracy and difficulty rating, we calculate the correlation coefficient between recognition accuracy and difficulty rating of performing a gesture for all five groups. The correlation coefficient of -0.45 shows medium correlation [21] between accuracy and difficulty rating, meaning that the higher accuracy a user achieves, the lower difficulty he/she is likely to rate performing a gesture.

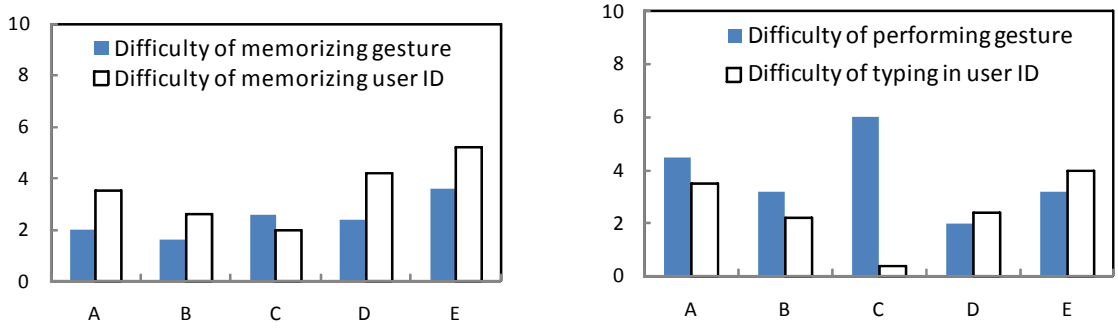


Figure 5: Survey result for difficulty of memorizing (left) and performing a gesture (right) for Group A to E

For Group D and E who receive both complexity constraint and the rejection procedure for ID gesture selection, the difficulty of performing ID gestures are perceived as similar to that of typing in a pre-composed textual ID.

5. CRITICAL AUTHENTICATION

Critical authentication aims at guarding privacy-sensitive data from unauthorized access. It is important to note that we do not expect gesture-based authentication to provide strict security but consider it as a convenient light-weight authentication method that can be combined with traditional methods. For example, if the device receives several false gestures, it can activate conventional password protection. We next explore whether uWave can recognize an owner-created gesture reliably while withstanding malicious forging, or *attack* as we referred to in this paper.

5.1 Objectives

We seek to answer the following questions through user studies.

- What tradeoffs between security and usability can the uWave-based solution achieve?
- How security can be jeopardized if the attacker sees the owner's gesture performance as gestures are much more visible than textual password entry?

5.2 Procedure

5.2.1 Participants

We recruit ten participants, three females and seven males, aged from 20 to 32. Nine of them are graduate students and one is undergraduate student. Their majors include Electrical Engineering, Computer Science, Psychology, Physics, Applied Mathematics, and Bio-engineering. We assign them to two five-person groups, F and G, in order to study the impact of visual disclosure.

5.2.2 Tasks of the Participants

In the study, the participants verify themselves with their password gestures and attempt to forge their group peers' password gestures. A participant is called the *owner* of his/her own password gestures but called *attacker* when he/she tries to forge the password gestures from others. The only difference between Groups F and G is that attackers in Group F do not see the owner performing the password gestures; attackers in Group G do see it through a video recording, which we call *visual disclosure*. For visual disclosure, the recording camera faces the front of the performers for all password gestures.

The study takes five days. On the first day, each participant selects two password gestures, each for one form of recognition feedback as explained later. In the following four days, the participant comes back for two tasks: 1) to perform their own password gestures for five times to verify themselves; 2) to forge the password gestures of other participants in the same group for five trials; once the attacker has the first successful attack, he/she will have five extra trials after that. One participant attacks a different victim on each day. As a result, each password gesture is attacked for at least 20 times by four participants. Figure 6 shows the gesture selections of Groups F and G.

It is important to note that we choose to allow only five trials in forging a password gesture by an attacker because the system can resort to a more reliable authentication method, such as conventional password, when several attempts have failed. Such a paradigm has already been widely used in other forms of authentication.

5.2.3 Two Forms of Authentication Feedback

We also study the effect of different forms of recognition feedback by providing the authentication results in two forms: success/fail and matching distance. Each participant has two password gestures. When the first password gesture is verified by the owner or attacked by an attacker, the authentication feedback is whether he/she succeeds or fails. The recognition result is based on a predetermined default threshold, calculated as a quarter of the base distance (defined in Section 3.1). The trials with success/fail feedback are used to generate the baseline performance with the fixed threshold. For the second password gesture, the feedback is the matching distance between the input gesture and the gesture template of the owner recorded at the first day of the study. With the matching distance feedback, the participants know whether they are getting closer or not. The trials with matching distance feedback are used for ROC analysis with various thresholds.

5.3 Authentication Results

Experiments with Group F demonstrate that uWave can achieve state-of-the-art false positive rate and false negative rate when the attacker does not see the target gesture. Not surprisingly, attackers in Group G encounters higher false positive rate due to visual disclosure. With a closer look into the results with matching distance feedback, however, it is possible to achieve both high usability and security for both Groups F and G if the rejection threshold can be set individually for each owner.

5.3.1 Baseline Performance with Default Threshold

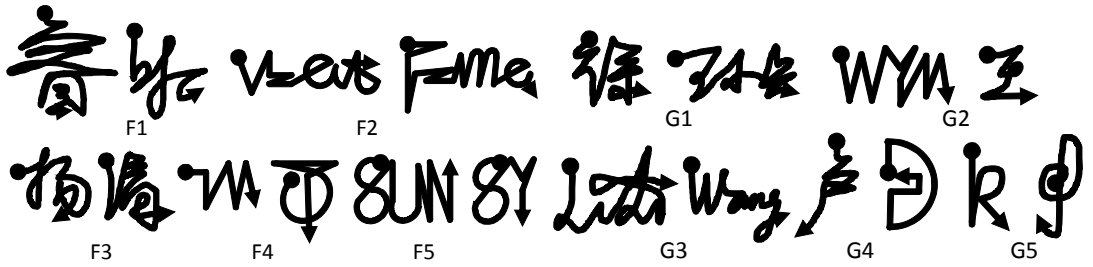
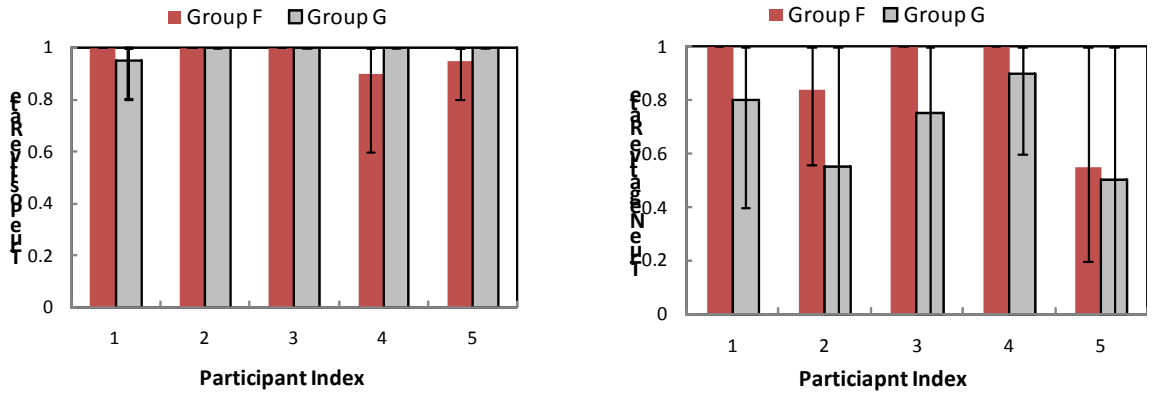


Figure 6: Password gestures for critical authentication (two for each participant)



(a) True positive rate **(b)True negative rate**

Figure 7: Baseline performance for each owner of Group F and G with success/fail feedback

The default threshold in trials with success/fail feedback is set as a quarter of the base distance. Figure 7 presents the results from this default threshold, where the true positive rate equals one minus false negative rate and the true negative rate equals one minus false positive rate. As in Figure 7 (a), results for trials with success/fail feedback show that uWave can correctly recognize the input gesture all the time for all participants except F5 and G1. F5 and G1 each have two and one false negatives, respectively. Therefore, the rejection threshold based on a quarter of the base distance leads to 98% and 99% average true positive rates for Groups F and G, respectively.

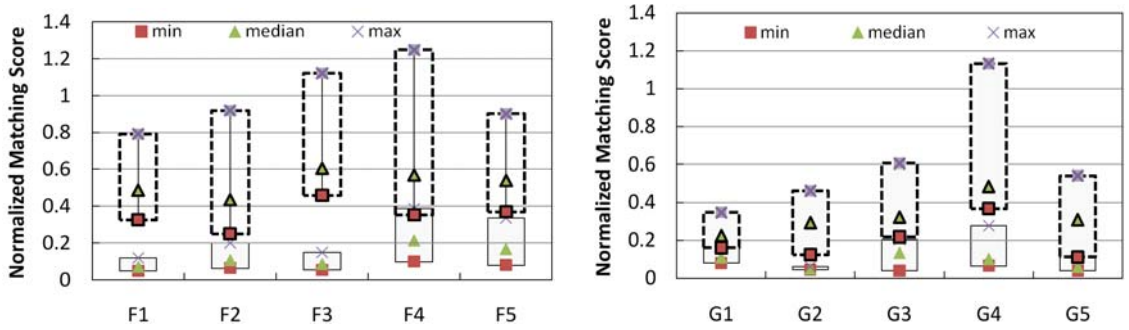
As to security as shown in Figure 7(b), all forged gestures are correctly rejected for all Group F participants except F2 and F5. uWave falsely accepts one to four of the faked gestures from other participants as they attack the password gestures of F2 and F5: F1 has one successful attack on F2 but fails to repeat it in the following five trials; F5 has the first success targeting F2 in the fourth trial and achieves four successful attacks among five additional trials. F1, F2, and F3 have two to four attack successes against F5. Overall, the same rejection threshold leads to 88% true negative rate.

With the same threshold, true negative rates are significantly lower in Group G in which attackers were given visual disclosure. The average is 70% for Group G versus 88% for Group F. It is important to note that their true positive rates are different too. Therefore, the difference between their true negative rates should not be interpreted as 88-70=18%, as we will see later in ROC analysis.

5.3.2 How close are the attackers?

For trials with matching distance feedback, Figure 9 shows the statistics of matching distances per participant in the form of box plots. The distances are normalized by the base distances of each password gesture. For Group F, the matching distances by attackers are always higher than those by the owner. It means if the threshold of rejection is carefully selected for each owner, it is possible to achieve 100% true positive rate and true negative rate for all owners except F4. If the proper threshold for an owner can be learned over time from multiple input samples, the performance can be significantly better.

For Group G in which visual disclosure is given to attackers, Figure 9(b) shows non-trivial difference between the matching distances by the owner and those by the attackers. Despite the low true negative rate with 0.25 as rejection threshold, matching distances by attackers are higher than those by the corresponding owner, similar to Group F. Hence it is still not trivial for attackers to forge a password gesture even with visual disclosure.



(a) Group F (b) Group G

Figure 9: Matching distance of the owners (boxes with solid outlines) and the attackers (boxes with dashed outlines)

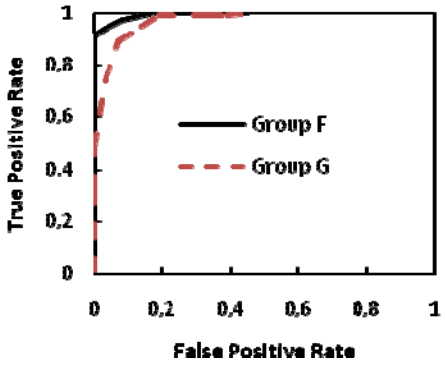


Figure 10: ROC curves of the uWave-based critical authentication

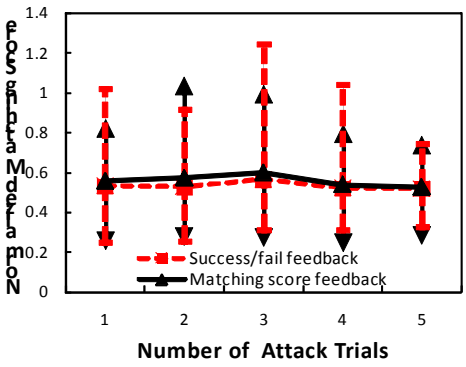


Figure 8: Normalized matching distances over multiple consecutive attack trials

5.3.3 ROC Analysis

Although it is possible to tune the rejection threshold for each user individually, an understanding of how uWave performs with a common threshold for all users is still important. To illustrate the tradeoffs between true positive and false positive in this case, **Figure 10** presents the receiver operating characteristics (ROC) curves for Groups F and G. We calculate the average true positive rates and false positive rates on all participants in each group by varying the rejection threshold from 0 to 0.5.

The ROC curve can help us decide a common threshold for all owners to achieve different tradeoff between false positive and true positive. A threshold between 0.15 and 0.2 will deliver nearly 95% true positive rate and below 2% false positive rate for Group F and 90% true positive and 5% false positive rate for Group G. Using the ROC curve, we can also estimate the equal error rate (when false positive rate and false negative rate is the same) as 3% for Group F and 10% for Group G.

5.3.4 Impact of Visual Disclosure

Not surprisingly, our study shows that visual disclosure increases false positives. As shown in **Figure 10**, under the same true positive rate, the false positive rate of Group G can be up to 10% higher than that of Group F. Such high false positive rate is likely to make the proposed authentication method useless, even for applications that do not require strict security.

5.3.5 Impact of Feedback

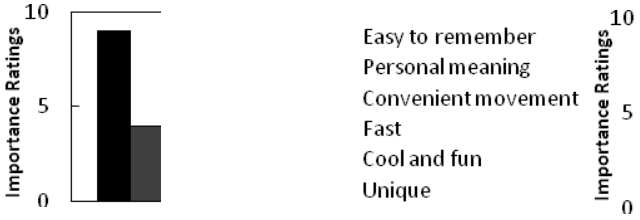
To explore the impact of different forms of feedback, we calculate the average matching distances of the attackers from the first trial through the fifth trial and present them in Figure 8. We make two observations. First, there is no clear trend in the attackers' performance as the number of trials increases. Second, there is no significant difference between the matching distances of success/fail feedback and those of matching distance feedback. We conjecture that the time series of acceleration is very complex and the space of exploration is simply too large to explore blindly, even with the feedbacks. As a result, even if the attackers know how close they are, it is still challenging for them to improve their attack.

6. DISCUSSION

We discuss some of our findings below in the context of future work.

6.1 Improving Critical Authentication

While our user evaluation shows that accelerometer-based authentication works well for non-critical authentication in terms of both usability and accuracy, it apparently cannot provide the strict security required by critical authentication in many applications when visual disclosure is unavoidable. However, with an equal error rate of 3% in the case of no visual disclosure, it is still promising for applications in which strict security is not necessary or it can be combined with other methods to achieve an even lower rate.



(a) Non-critical authentication (b) Critical authentication

Figure 11: Importance ratings of factors in gesture selection

As we show in Section 5.3, visual disclosure can potentially render the authentication method useless even for applications that do not need strict security. Therefore, visual concealment is needed. While it is difficult to prevent others from seeing one perform the gesture, one may be able to hide the starting and end points of the password gesture. This can be easily implemented on many platforms. For example, our Wii remote-based implementation requires the user to hold a button on the remote while performing a gesture. Since it is difficult for the attackers to clearly see whether the user has pressed the button or not, the user can add spurious movement before pressing the button or after releasing the button to hide the real gesture. We also suggest employing 3D movements in the password gesture in order to make it more challenging to forge.

6.2 Choice of Gestures

It is interesting to note how our participants compose their ID and password gestures. First, the selected gestures are very symbolic, such as regular shapes, letters, and characters in the native languages of the participants. Unlike speech or handwriting for which we are familiar with a well defined vocabulary, gestures are not employed in our everyday life for human to computer interaction so that lack of a defined vocabulary commonly accepted by users. Therefore, gestures based on well-known concepts and symbols are easier to memorize as well as to perform consistently. Second, the selected gestures often carry personal meanings. For example, some of them are the name initials of the participants. Such choice provides an easy solution for the uniqueness of gestures that can be easily memorized. Third, not surprisingly, password gestures for critical authentication from Groups F and G are significantly and consistently more complicated than ID gestures for non-critical authentication from Groups A to E.

For non-critical authentication, we note that uWave works well for both collective and individual procedures of ID gesture selection. That it works well for collectively selected gestures implicates that uWave distinguishes gestures in a similar way human users do; that it works well for gestures selected under uWave supervision implicates that uWave is effective in guiding users for rapidly selecting proper gestures without knowledge of others' gestures.

To further understand gesture selection, we ask the participants to rate the importance of several factors in their gesture selection in the survey. These factors include "easy to remember", "having personal meaning", "convenient for hand and arm movement", "fast to perform", "cool and fun to perform", and "likely to be unique". The average ratings are shown in Figure 11(a). Not surprisingly, "easy to remember", "convenient for movement" and "fast to perform" are the most important three factors for non-critical authentications. All three factors are concerned with usability, memorizing and performing the gesture.

In contrast, the three most important factors for critical authentication are "unique", "easy to remember", and "personal meaning." While participants still care about "easy to remember", they consider security as more important than difficulty of performing gestures. "Unique" is rated significantly higher than in non-critical authentication, indicating that the participants consider uniqueness lead to better security. In addition, "personal meaning" also receives considerably higher rating than in non-critical authentication. When answering the open ended questions about their gesture selection, the participants in critical authentication indicate personal meanings help them remember rather complicated gestures. In contrast, participants in non-critical authentication select simpler gestures and do not need personal meanings to help them remember the gestures.

As mentioned in Section 5.3, we also observe gesture selection can have a great impact on the tradeoff between security and usability for critical authentication. Our observations suggest that sharp movement changes in gestures create fine features in the time series of acceleration and therefore can make it more challenging to forge.

7. CONCLUSIONS

In this work, we investigate the feasibility and usability of gesture-based user authentication using uWave, a gesture recognition system based on a tri-axis accelerometer. For non-critical authentication, uWave recognizes the user from a small group of possible users; for critical authentication, uWave verifies the claimed user identity. We report an extensive evaluation of gesture-based user authentication with a comprehensive set of user studies.

For non-critical authentication, we draw the following conclusions from our user studies.

- UWave achieves an average accuracy of 98% in recognizing users based on user-created ID gestures even over multiple weeks.
- Our participants rate the difficulty of memorizing and performing ID gestures as no more than that of our pre-composed simple textual IDs, given proper gesture selection constraints.
- Gesture selection constraints have a significant impact on the accuracy of accelerometer-based authentication. Gesture complexity constraints and the rejection procedure can improve accuracy significantly.
- Users' evaluation of the difficulty of performing ID gestures is related to the accuracy they achieve. The higher accuracy the less difficult they tend to rate. Therefore, potential techniques intended for accuracy improvement are also likely to make users feel easier to use gestures for authentication.

For critical authentication, we draw the following conclusions.

- Without visual disclosure, an equal error rate 3% is achieved by uWave with a single training sample, compared to that reported in [13] (4%) with substantially more training data.
- Visual disclosure increases the false positive rate by up to 10%, given the same true positive rate. It also increases the equal error rate to 10%.
- Comparison between two forms of feedback shows that knowing the matching distance does not help the user achieve higher accuracy.

In summary, our research has demonstrated that accelerometer-based gesture recognition can provide feasible and usable solution for non-critical user authentication. For critical authentication, uWave achieves the state-of-the-art performance without visual disclosure. With 3% equal error rate, it can be useful when strict security is not expected. However, we show that visual disclosure can potentially increase the equal error rate to 10%, making the authentication method useless even for non-strict security. There is a need for future research to cope with visual disclosure.

With the proliferation of low power, low cost accelerometers, we believe accelerometer and gesture-based user authentication has the potential to enable personalized services on resource-constrained mobile devices. The work reported here, nevertheless, is the first step toward this goal. We believe further feature analysis of acceleration from physical manipulation, inspiration from more sophisticated solutions in handwritten signature verification, and adaptive solutions to adjust the rejection threshold can help achieve more effective and usable gesture-based authentication.

8. ACKNOWLEDGEMENTS

The work is supported in part by NSF awards IIS/HCC 0713249 and CNS/CSR-EHS 0720825 and by a gift from Motorola. The authors would like to thank Crysta Metcalf and Elaine Huang from Motorola for their help with our user study design. The authors would also like to thank the participants in the user studies and thank the anonymous reviewers whose comments helped improve the final version of this paper.

9. REFERENCES

[1] J. Liu, Z. Wang, L. Zhong, J. Wickramasuriya, and V. Vasudevan, "uWave: Accelerometer-based Personalized Gesture Recognition and Its Applications," in *Proc. IEEE Int. Conf. Pervasive Computing and Communication (PerCom)*, 2009.

[2] F. G. Hofmann, P. Heyer, and G. Hommel, "Velocity Profile Based Recognition of Dynamic Gestures with Discrete Hidden Markov Models.," in *Proc. Int. Wrkshp. Gesture and Sign Language in Human-Computer Interaction*, 1997.

[3] I. J. Jang and W. B. Park, "Signal Processing of the Accelerometer for Gesture Awareness on Handheld Devices," in *Proc. IEEE Int. Wkshp. Robot and Human Interactive Communication*, W. B. Park, Ed., 2003, pp. 139-144.

[4] J. Kela, P. Korpipää, J. Mäntyjärvi, S. Kallio, G. Savino, L. Jozzo, and D. Marca, "Accelerometer-based gesture control for a design environment," *Personal Ubiquitous Computing*, vol. 10, pp. 285-299, 2006.

[5] B. D. Payne and W. K. Edwards, "A Brief Introduction to Usable Security," *IEEE Internet Computing*, vol. 12, pp. 13-21, 2008.

[6] D. Maltoni, Handbook of fingerprint recognition: Springer, 2003.

[7] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, vol. 35, pp. 399-458, 2003.

[8] R. P. Wildes, "Iris Recognition: an Emerging Biometric Technology," *Proc. IEEE*, vol. 85, pp. 1348-1363, 1997.

[9] J. P. Campbell, Jr., "Speaker Recognition: a Tutorial," *Proc. of the IEEE*, vol. 85, pp. 1437-1462, 1997.

[10] D. Impedovo and G. Pirlo, "Automatic signature verification: the state of the art," *IEEE Trans. on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 38, pp. 609-635, 2008.

[11] F. Okumura, A. Kubota, Y. Hatori, K. Matsuo, M. Hashimoto, and A. Koike, "A Study on Biometric Authentication Based on Arm Sweep Action with Acceleration Sensor," in *Proc. Int. Symp. Intelligent Signal Processing and Communications*, 2006.

[12] E. Farella, S. O'Modhrain, L. Benini, and B. Riccó, "Gesture Signature for Ambient Intelligence Applications: A Feasibility Study," in *Proc. Int. Conf. Pervasive Computing (Pervasive)*, 2006.

[13] K. Matsuo, F. Okumura, M. Hashimoto, S. Sakazawa, and Y. Hatori, "Arm Swing Identification Method with Template Update for Long Term Stability," in *Proc. Int. Biometrics*, 2007.

[14] J. Mantyjarvi, M. Lindholm, E. Vildjiounaite, S. M. Makela, and H. A. Ailisto, "Identifying Users of Portable Devices from Gait Pattern with Accelerometers," in *Proc. of IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, 2005, pp. ii/973-ii/976 Vol. 2.

[15] K. Hinckley, "Synchronous Gestures for Multiple Persons and Computers," in *Proc. ACM Symp. User Interface Software and Technology (UIST)*, 2003.

[16] R. Mayrhofer and H. Gellersen, "Shake Well Before Use: Authentication Based on Accelerometer Data," in *Proc. Int. Conf. Pervasive Computing (Pervasive)*, 2007.

[17] S. N. Patel, J. S. Pierce, and G. D. Abowd, "A Gesture-based Authentication Scheme for Untrusted Public Terminals," in *Proc. ACM Symp. on User Interface Software and Technology (UIST)*, 2004.

[18] D. Kirovski, M. Sinclair, and D. Wilson, "The Martini Synch: Joint Fuzzy Hashing Via Error Correction," in *Proc. European Wrkshp. Security and Privacy in Ad-hoc and Sensor Networks*, 2007.

[19] L. E. Holmquist, F. Mattern, B. Schiele, P. Alahuhta, M. Beigl, and H.-W. Gellersen, "Smart-Its Friends: A Technique for Users to Easily Establish Connections between Smart Artefacts," in *Proc. Int. Conf. Ubiquitous Computing*. Atlanta, Georgia: Springer-Verlag, 2001.

[20] D. L. Clason and T. J. Dormody, "Analyzing Data Measured by Individual Likert-Type Items," *Journal of Agricultural Education*, vol. 35, No. 4, pp. 31-35, 1994.

[21] J. Cohen, P. Cohen, S. G. West, and L. S. Aiken, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*: L. Erlbaum Associates Mahwah, NJ, 2003.