BroAuth: Evaluating Different Levels of Visual Feedback for 3D Gesture-Based Authentication

Max-Emanuel Maurer, Rainer Waxenberger, Doris Hausen
University of Munich
Media Informatics Group
80333 München, Germany
max.maurer@ifi.lmu.de, waxenberger@cip.ifi.lmu.de, doris.hausen@ifi.lmu.de

ABSTRACT

Using digital gadgets we authenticate ourselves regularly. Usually authentication relies on standard PIN or password but novel input hardware facilitates new authentication techniques. In this work we present an authentication mechanism based on body movements captured by a depth sensor. This idea is motivated by the cultural body movements used as welcoming gestures, especially by gang members (secret handshakes). Our authentication technique 'BroAuth' lets the user interact with a virtual partner to perform password input. This is done through touching target zones on the own body and on the body of a virtual partner.

In this paper we focus on evaluating usability and security of onscreen feedback for such a system. Three different types of feedback were tested during the input process: Text-only (1D), abstract user representation (2D) and a virtual avatar (live 3D). The most detailed but most insecure 3D feedback performed much worse than the abstract input modalities. Input times and user opinions show that an abstract 2D representation is the best tradeoff between usability and security for such a system.

Categories and Subject Descriptors

D.4.6 [**Software**]: Security and Protection

General Terms

Design, Security, Experimentation

1. INTRODUCTION

Our daily live is getting more and more digital. Mobile phones, computers, ATMs and gaming consoles are all part of ubiquitous networks and store data and user profiles. In all those situations people have to prove themselves legitimate to the devices to perform specific actions. By now a vast number of authentication mechanisms have been proposed, but PIN and password are still applied to nearly every new piece of technology. However, the concept of graphical passwords is getting more popular [6].

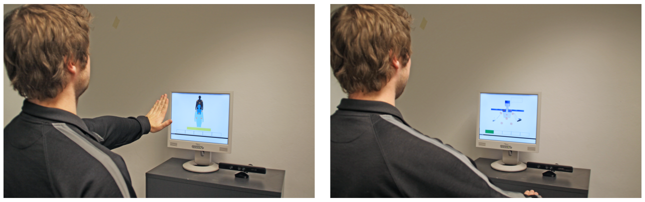

Figure 1: With the BroAuth prototype participants authenticate by touching certain body zones of a virtual partner or themselves. We tested different feedback (left: 2D, right: 3D; compare fig. 3).

With every new input technology, new authentication mechanisms become possible. Depth sensors that are able to provide live 3D data of the space lying in front of them are now inexpensively available (e.g. Microsoft Kinect). In this work we present a new authentication approach and especially address two questions: (1) How to design onscreen feedback when authenticating with a depth sensor? and (2) how much feedback is best when interacting with such a system in regards to usability and security?

With 'BroAuth' the user authenticates by performing a series of gesture inputs – i.e. touching different zones on her own or a virtual partner's body (see figure 1). Body movements are also used in real life to transport certain types of information. Cultural behavior is a topic used in computer science research in different domains [4]. Welcoming rituals are cultural body movements with meanings that are known for centuries. But in other social communities body movements – like secret handshakes – even present some form of authentication. Gang members for example use certain patterns of movement salutatory to prove their membership. Although it is obvious that such a system may never be perfectly secure it could benefit from the motor memory effect [7].

Since the virtual partner is not physically present the appropriate amount of visual feedback is an important research issue. For security and usability reasons the best tradeoff between providing feedback to the user and hiding the input has to be found. Besides presenting the new authentication concept, insights on visual feedback in this context are our main contribution.

AVI '12, May 21-25, 2012, Capri Island, Italy
Copyright 2012 ACM 978-1-4503-1287-5/12/05... $10.00

2. RELATED WORK

In an attempt to create alternatives for PIN and password authentication, researchers have designed different authentication methods that promise improved memorability.

Weiss and De Luca discovered that many users memorize their PINs by remembering the shape resulting from the order of their PIN's single digits on the number pad [7]. They developed PassShape, that replaces PIN numbers with shapes consisting of directional two-dimensional strokes. The PassShape approach wants to achieve increased memorability by using shapes as visual password mnemonics and by stimulating the motor memory as an effect of drawing the shape repetitively in the same order. The evolution of PassShapes named EyePassShapes combines the PassShape concept with the input method of eye gestures [2]. As one part of the EyePassShapes evaluation process several visual background designs have been tested to determine the appropriate visual support for the complex input procedure.

Chong and Marsden [1] further exploit the use of gestures for authentication. They use several 3D movements as password elements. To generate a gesture password a sequence of predefined movements has to be performed. The built-in accelerometers of mobile phones are used to sense the movements of users. The gesture password approach was motivated by the motor memory effect to improve memorability in comparison to common PIN entry.

3. OUR CONCEPT AND PROTOTYPE

BroAuth makes use of body movements as a mean for authentication. The idea is adopted from two people welcoming each other using body movements and interaction instead of a simple handshake – e.g. two gang members. In our concept the password consists of a set of predefined moves that have to be performed by the user. The computer provides a virtual partner for the authentication process.

In the physical world such a process can get very complex. The number of different limbs that take part and the granularity of the interaction may vary greatly. A 'fist bump' and a 'high five' may seem similar at first but are totally different gestures. For our concept we reduced the set of possible interactions to a minimum as we did not focus on finding an optimum tracking solution for this kind of authentication, instead we investigated the level of visual feedback that would be needed.

BroAuth allows for eight different gesture inputs by touching 'target zones' at either the own body or the body of a virtual partner (see figure 2). To enable the user to find and interact with the invisible virtual partner, we use a virtual 'clone' of the user herself (in terms of height, shoulder width etc.). The clone is facing the user and is placed always right in front of the user at arm's length.

3.1 Hardware and Software Realization

The prototype is based on a visual 3D tracking system to recognize users and capture their three dimensional movements. The tracking system was realized using the Microsoft Kinect depth sensor and compatible software components. The main software component is the OpenNI framework that enables direct access to the data produced by the depth sensor. The OpenNI API provides functions for user recognition and body tracking that can be implemented by NITE middleware. Among others NITE provides a specific algorithm that fits a special skeleton model on the registered user. This skeleton model consists of 15 nodes (skeleton joints), each corresponding to a certain point of the human body. Three-dimensional movements and gestures can be determined by the position of the single skeleton joints.

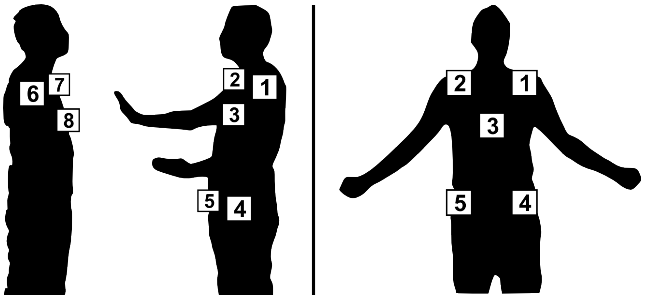

Figure 2: Left: The eight different touch zones that can be used for password entry. Five zones for the user herself and three at the virtual partner. Right: The five personal touch zones of the user herself.

The interaction model of the BroAuth prototype uses eight interaction points (see figure 2) derived from the current position of the OpenNI/NITE skeleton joints. The implementation in C# registers an interaction when the user places one of his hands in one of the defined areas. A sequence of registered interactions can be considered as a password.

3.2 Password Space

Passwords for BroAuth may consist of an arbitrary combination of eight touch zones but the same touch zone may not be immediately repeated. This restriction was made to allow a faster input with no dwell times or similar measures. To get a password space that has roughly the same size as a 4-digit PIN we used five touches per password during our study. This resulted in a total of 19,208 input combinations (cf. 10,000 for PIN).

3.3 Security

Besides input speed and password space security against attacks is important. The basic BroAuth concept is so far not protected against the common attacks on password input – e.g. shoulder-surfing. But one advantage of the concept lies in the capability of human motor memory. Remembering certain body movements usually requires training and an unskilled attacker may be unable to memorize and replay a password while it can be perfectly remembered by the user. Moreover, once trained movements can be retained over a long period [5]. The concept as it is presented here, should only be used in less secure or public contexts (e.g. gaming consoles, domestic use). We did not focus on optimizing the security in this work as we investigated visual feedback of such a system.

4. ONSCREEN FEEDBACK

Independent of the authentication mechanism that is used it is always crucial to provide feedback for each input and authentication state. For a gesture based password this feedback might be different from what is common use today.

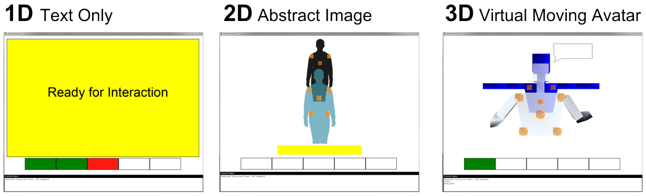

Figure 3: The three different visual feedbacks used in the study. 1D text-only, 2D abstract avatar and 3D moving avatar. All three feedback types use the same progress indicator.

How much of the aggregated data needs to be displayed for the user to make correct inputs? In which way should this data be presented? This is always a tradeoff between usability and security. We take a closer look on this and compare different methods of visual feedback. We conducted a focus group to narrow the space of possible onscreen feedback and finally implemented a prototype with three different types of visual feedback.

4.1 Focus Group

To generate feedback concepts for BroAuth we conducted a focus group with five participants (two female). The group discussed four main topics: visualization, feedback, user guidance and security. Finally we had the participants draw their own concepts of such a visualization. For visualizing the input data users wanted a rather abstract representation concept. They were undecided between a virtual 3D avatar representing the user and an abstract 2D representation of the user and the virtual partner.

4.2 1D, 2D and 3D feedback

Consequently we designed three concepts with increasing visual feedback and thus decreasing security (see figure 3).

- **Text-only or 1D feedback:** simple text commands and feedback.
- **2D:** abstract 2D images of user and partner. All interaction points are displayed and highlighted all together whenever an interaction happens.
- **3D:** three dimensional avatars for user and partner. The users avatar moves according to his own arm movements in realtime. Touch zones are displayed in 3D and highlighted all together whenever an input occurs.

5. EVALUATION

To evaluate our concept and the three different user feedback designs, we conducted a user study to find out about the performance of an untrained user when interacting with the system and more importantly, which feedback design would perform best.

5.1 Hypotheses

We had five hypotheses for the study: **H1:** The more detailed the visual feedback is the stronger users will focus on it. **H2:** With increasing visual feedback the input time will increase. **H3:** With increasing visual feedback the error rate will drop. **H4:** The more detailed the visual feedback is, the more usable it will be rated. **H5:** Experienced security will drop with increasing visual feedback.

Table 1: The avarage input times and error means and their standard deviations for the three different types of visual feedback.

	Time (s/task)		Errors (1/task)	
Cond.	**Avg**	**SD**	**Avg**	**SD**
1D	12,50	11,30	0,72	1,34
2D	12,34	6,66	0,68	1,27
3D	15,75	12,50	0,93	1,44
Overall	13,53	10,53	0,78	1,35

5.2 Methodology

Our study setup consisted of a Kinect sensor and a computer monitor displaying the visual feedback. Participants had to learn and repeat different movement patterns while using our system with different feedback modalities. We had a within-subjects design with each participant using all three types of visual feedback. We fully balanced the order of the appearing feedback over the participants. For each feedback modality we created a 'teaching mode' that instructed the participants, which body part had to be touched next. Participants were trained to a randomly generated password five times using the teaching mode and afterwards had to input their password again five times using the same type of visual feedback.

As dependent variables we measured the time for one interaction sequence and the number of input errors. Participants had to calibrate themselves for the tracking to work and used a start gesture (clapping over the head) to start each input sequence. This time was not measured as being part of the interaction. We also had an additional questionnaire to rate each type of visual feedback.

5.3 Participants

We had twelve participants for the study (three female) and most of them were computer science students. In average participants were 24 years old (SD 3.3). One participant was left handed. Nine participants had absolutely no prior knowledge of the Kinect system. Half of them had used other motion based systems (e.g. Nintendo Wii).

5.4 Results

Our results are based on two different types of data. Firstly we measured quantitative data from the users inputs: Input took on average 13.5 seconds (see table 1). 1D (avg 12.5s) and 2D (avg 12.3s) performed nearly equal while 3D (avg 15.8s) had a much higher average. Although this difference looks large, due to the high standard deviations no significant effect could be found (ANOVA with Huynh-Feldt correction: $F_{1.806} = 2.0; p = .143$). So we can only see a tendency towards the second hypothesis here.

Each user made on average 0.78 input errors per round. Again 2D and 1D performed slightly better than the 3D visualization but the changes were not significant. This means we unable to confirm our third hypothesis.

Besides this data we also asked participants about their experience with the different types of visual feedback. Most of the questions were rated from "1 strongly disagree" to "5 strongly agree".

For our first hypothesis we asked the participants whether they had been “closely focussed on the visual feedback”. Participants stated to be significantly more focussed on the 3D (Mdn 5) feedback than on the 2D (Mdn 4.5) and 1D (Mdn 3.5) feedback. A Friedman-Test showed an overall significant difference for the answers ($Z = 16.60; p < .001$). Using Wilcoxon as a Post-hoc test, all three levels showed significant differences.

We also asked people whether they found the task was not physically demanding. Ten people strongly agreed (Mdn 5). All people agreed or strongly agreed that it was fun using the authentication mechanism (Mdn 5).

For H4 and H5 we asked people about their assessment of usability and security of the visual feedback. Here we used a five-point Likert scale from “-2 too little” to “+2 too much”. Concerning usability the 3D representation as expected was rated best (avg 0,0; SD 0.4) followed by 2D (avg -0,17; SD 0.4) and 1D (avg -0,75; SD 0.8). Again these differences are significant ($Z = 6.87; p = .32$) thus confirming H4. Post-hoc analysis this time showed a significant effect only for 1D compared to 2D and 3D. This shows that participants experience the more accurate feedback as more usable but using 2D or 3D does not make a real difference to them – regarding this the focus group was already undecided. When combining these findings with the measured input times it is obvious that there is no necessity for a 3D feedback and even 1D feedback seems still to be a valid option.

We also asked the users for their perceived security and did not find any significant differences between 3D (avg 0.0; SD 1.4), 2D (avg -0.17; SD 0.7) and 1D (avg -0.08; SD 0.8). This result of 3D still being rated best was very unexpected since the visual feedback clearly shows a detailed motion of the users body. An attacker just recording the screen would be able to steal a password with this type of visual feedback. We suppose that participants may not have been aware of this and may have judged the security of the whole situation. H5 can hence not be confirmed.

In the end we asked the participants to sort the feedback types by personal preference. 2D was preferred by most participants (nine participants) followed by 1D (two participants) and 3D (one participant). This clearly shows which type of visual feedback was most pleasing.

6. LIMITATIONS

Looking at security as well as input speed our system – using whatever type of visual feedback – is not able to outperform common authentication measures – e.g. PIN. As mentioned earlier this was not the main purpose of this work. Instead we consider this as a first step in using body tracking in 3D as authentication and therefore focused on the visual feedback. Besides typical factors like security and speed our system may have advantages through the human motor memory. With the user being able to reproduce a trained gesture quite exact [3] a more restrictive input process including a time factor might be able to raise security and exclude attackers from successfully authenticating. The motor memory itself might also prove more powerful for learning and persistently retaining gesture passwords in mind. This both has to be validated in future work.

The error rate we had in the study might also be a problem in a real world setting. Many factors contributed to the error rate. We were mostly interested in real mistakes users made – because they did not remember the input pattern correctly. But we also observed that a large number of user slips occurred as users accidentally activated target points – e.g. with the non primary hand that was out of their focus.

For our first hypothesis we only relied on the users self reporting in the questionnaire. A real eye tracking mechanism would have been much better to determine how much users spent looking on the screen.

7. CONCLUSIONS AND FUTURE WORK

In this work we presented a new concept for authentication using a depth sensor and body movements in 3D by interacting with a virtual partner. Furthermore we evaluated which kind of onscreen feedback would be best suitable for such an input technique in terms of usability and security. Users found interactions to be fun and not physically exhausting. In terms of interaction time, errors and security the prototype performed similar to other proposed authentication approaches but was not able to outperform the standard means of authentication. Possible advantages of the motor memory were not yet part of the evaluation.

We also evaluated three different types of onscreen feedback for the input process. In terms of onscreen feedback an abstract 2D view was overall the best option for visualizing user input. Input times, error rates and user feedback were overall best here. Even simple text output (1D) performed better than the least secure live 3D view did. Only for usability the 2D and 3D feedback was rated significantly better than the 1D feedback. A security assessment by the participants did not show any significant difference.

For future work it is of great importance to perform research on how the human motor memory can contribute to the concept. What about password memorability in this context? And how can timing contraints in the password be used as an additional security layer?

8. REFERENCES

[1] M. Chong and G. Marsden. Exploring the use of discrete gestures for authentication. In *INTERACT*, 2009.

[2] A. De Luca, M. Denzel, and H. Hussmann. Look into my eyes!: Can you guess my password? In *SOUPS*, 2009.

[3] J. Krakauer and R. Shadmehr. Consolidation of motor memory. In *Trends in Neurosciences*, 2006.

[4] M. Rehm, N. Bee, and E. André. Wave like an egyptian: accelerometer based gesture recognition for culture specific interactions. In *BCS-HCI*, 2008.

[5] R. Shadmehr and T. Brashers-Krug. Functional stages in the formation of human long-term motor memory. *The Journal of Neuroscience*, 1997.

[6] X. Suo, Y. Zhu, and G. S. Owen. Graphical passwords: A survey. In *ACSAC*, 2005.

[7] R. Weiss and A. De Luca. Passshapes: utilizing stroke based authentication to increase password memorability. In *NordiCHI*, 2008.