Spinlock: A Single-Cue Haptic and Audio PIN Input Technique for Authentication

Andrea Bianchi[1], Ian Oakley[2], and Dong Soo Kwon[1]

[1] KAIST, Daejeon, Korea
[2] Madera ITI, University of Madeira, Funchal, Portugal
andrea@kaist.ac.kr, ian@uma.pt, kwonds@kaist.ac.kr

Abstract. Authentication in public spaces is inherently exposed to observation attacks in which passwords are stolen by the simple act of watching the data input process. Addressing this issue are systems that secure authentication input via PINs or passwords that rely on sets of relatively unobservable tactile or audio cues. However, although secure, such systems typically invoke high levels of cognitive load in their users which is instantiated in lengthy authentication times and high error rates and most likely due to significant cognitive demands in terms of processing, mapping or recalling non visual information. To address this issue this paper introduces Spinlock, a novel authentication technique based on repeated presentation, recognition and enumeration of a single, simple invisible cue (audio or haptic), rather than a set of structured stimuli. This approach maintains the security but avoids the complexity of previous systems. A prototype illustrating this concept is described as well as a study comparing modalities and gauging overall levels of performance, usability and security. The results show that authentication with Spinlock is faster and less error prone than previous non-visual systems, while maintaining a similar security level. Limitations and future work are discussed.

Keywords: Authentication, haptic and audio PIN, mobile.

1 Introduction

Users' interaction with PIN-entry interfaces situated in public spaces is inherently observable by third parties. While this is acceptable in many situations, such as while interacting with information kiosks, it is problematic during confidential interactions such as PIN entry at bank ATMs or public password entry on mobile devices. In these cases, the observable nature of the input device becomes a weakness that can be subjected to observation attacks, both in person (a technique known as shoulder surfing) and via appropriately positioned video recording equipment (a camera attack). These risks are significant – ATM fraud in the USA is estimated to run to 60 million USD per year [1] – and have been comprehensively discussed in the research community [2].

In order to create observation resistant data entry techniques, recent research has explored the use of invisible cues, such as audio or haptics, as an alternative input/output method to support PIN entry in public terminals [e.g. 3, 4]. Fundamentally, the argument underlying this work is that the highly physical or

E.W. Cooper et al. (Eds.): HAID 2011, LNCS 6851, pp. 81–90, 2011.

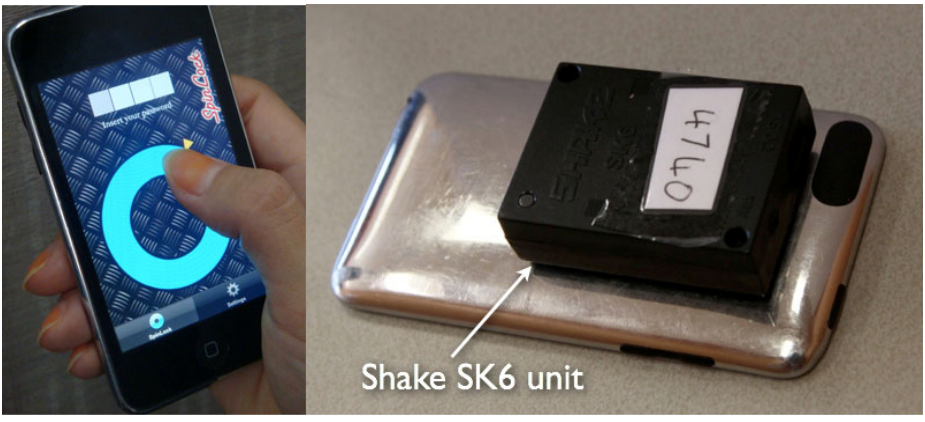

Fig. 1. Inserting a PIN in the Spinlock application (left) and the Spinlock hardware setup (right) - an Apple Ipod Touch and a Shake SK6 unit for generating vibration stimuli

proximate nature of invisible cues (touch or audio through headphones) makes it more difficult for observers to intercept key information both in person and via recording equipment –that observing haptic/audio information is more challenging than observing traditional PIN entry activities, such as key presses.

Current research supports this suggestion [e.g. 3, 4]. However, the use of these cues is not without its limitations. Most importantly, while harder for a third party to observe, a set of invisible cues (often in the form of tactons [5]) is also more challenging for a user to accurately perceive, process and interpret. Secure haptic data entry tasks, for instance, have typically resulted in high levels of cognitive load expressed empirically through lengthy task completion times and high error rates [3, 4]. The work described in this paper aims to retain the observation-resistant property of haptic and audio cues - the threat model considered in this work is malicious observation of PIN entry in person and via recording equipment in a public space. However, this paper aims to mitigate the cognitive effort required to interpret such non-traditional password cues. It does this by presenting the design of Spinlock, a system based on the repeated display of a single, simple and easy to recognize cue, rather than a set of structured invisible stimuli [e.g., 3]. It also explicitly compares audio with haptic cues and presents a discussion of the differences observed.

The remainder of this paper is structured as follows: a literature review; a description of the conceptual structure of the security system and the details of a prototype for mobile phones that instantiates it; a user study incorporating usability and security evaluations; a discussion of the results and speculations for future work.

2 Related Work

Researchers have explored a wide range of haptic and audio techniques for PIN entry. Most of early work in this area used a multi-modal approach, combining the rich visual modality of graphical or textual passwords with haptic or audible cues. For example, in early work on this topic Malek et al. [6] described haptic passwords that

used pressure-based input as a hidden channel to obfuscate entry of an otherwise graphical password. In this system users drew a password composed of lines connecting points on a grid, and the pressure applied during drawing was used as supplementary information to compose the password.

More recently, both Sasamoto et al [4] and de Luca et al [7] described data entry techniques based on the combination of observable visual input modified via a users perception of unobservable tactile cues in the form of directional strokes applied to the skin or the vibrations of a mobile device. Although promising, these approaches require users invest significant cognitive resources in order to map known actions to sensed haptic stimuli, or rely on the user's perception and recognition of hidden haptic cues in order to transform their observable input. Such mental mappings are not trivial and lead to lengthy authentication times and high error rates: for instance in Sasamoto's Undercover system median task completion times are reported to be 25-45 seconds, with error rates of between 26%-52% [4].

In contrast to this multi-modal approach, Bianchi et al. [3] proposed a uni-modal haptic password based on selecting a sequence of tactons in much the same way as numbers are selected on a regular keypad. To secure against observation, the tactons were randomized over the keys between selections. The task in this system is simply to recognize and select haptic cues and the authors argue this simplicity should result in lower levels of cognitive load (and correspondingly improved task completion times and error rates) when compared to multi-modal approaches. Evaluations of a number of system variations, including an audio entry systems that works analogously to the haptic version [8, 9], support this claim (authentication in less than 20 seconds, with 7% mean error rate) as does highly related work by Kuber and Yu [10], in which a similar concept is instantiated based on spatially varying cues rendered on Braille cells explored by the fingertips. However, a disadvantage of such systems is that they require users to accurately select particular haptic cues from a stimulus set, a challenging task when sets exceed 3 or 4 items in size [e.g. 3, 8, 9, 11]. Issues of learning and retention of tactons are also poorly understood - from the perspective of human cognitive limits, it is currently unclear how scalable and reliable the concept of a purely haptic password really is [11]. These issues place doubts on the viability of these recognition-based approaches.

On the other hand, work on audio authentication has typically focused on identity recognition and used speech as an auxiliary input modality in combination with other biometric techniques (e.g., lip sync, fingerprints, face recognition) [12, 13]. Although these systems do provide stronger multi-factor authentication based on orthogonal data sources, they do not attempt to offer a direct solution to the observation attack; voice can be easily recorded in public spaces using directional microphones and such systems can be sensitive to replay attacks utilizing playback of such data.

The work in this paper addresses these issues. Its contribution is the design of a PIN entry system that relies on simple uni-modal haptic or audio cues, but that does not require users to learn or distinguish among a large set of distinct stimuli, nor use the audio or haptic modality as a compliment to other input. It achieves this via the rapid, repeated display of a single, brief and distinctive cue in response to user input. By counting the number of displayed cues (either haptic or audio), users can enter structured data. This design seeks to retain the advantages of non-visual uni-modal PIN entry while sidestepping issues of learning and recognizing a stimulus set.

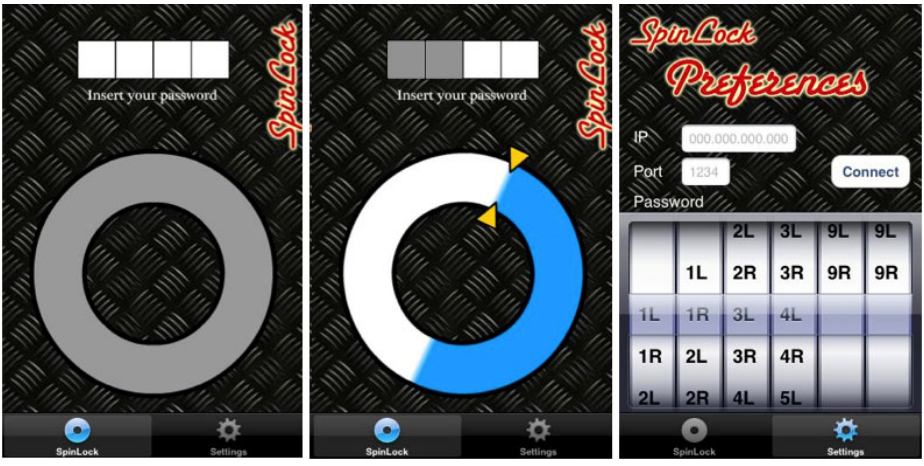

Fig. 2. The Spinlock graphical user interface: whilst idle (left), during the user interaction with two PIN items entered (center) and the settings screen showing user password (right)

3 Design and Implementation of the Spinlock PIN Entry System

The Spinlock prototype is based on the dial-lock of a safe. In such systems, PINs are composed of a sequence of numbers and a direction of motion (clockwise/right or anti-clockwise/left), which must alternate. For example, in a dial marked with 10 numbers, a four-item PIN could take the form of the following rotations: 2-left, 8-right, 5-left, and 7-right. Spinlock is based on a similar interaction with two key differences. Firstly, the requirement to alternate directions is removed (via the provision of widget deselection, an additional input delimiter). Secondly, rather than moving to a number marked on a dial, users count the number of audio or haptic cues delivered during their input. Upon termination, the direction of their motion and the number of cues they experienced constitute the PIN item sent to the system.

For example, to enter the password listed above, users would input leftward rotation until two audio or haptic cues were experienced, followed by rightward motion for a count of eight cues, leftward for five and finally rightward for seven. Although this password features alternating directions, this is not a requirement for the Spinlock system - input can also be delimited by deselection of the control widget.

In order to remain resistant to observation the spatial distance users must travel between cue presentations is randomized (among 7 possible distances, 12º apart from each other, ranging from 36º to 120º) after every cue. The goal of this manipulation is to increase the resistance of the system to attack via visual observation. It decouples the distance that the Spinlock dial is rotated from a direct correspondence with the data that is input.

To explore the validity of this design, Spinlock was implemented for the Apple iPhone and iPod Touch devices (Figures 1 and 2). The touch screen was used for input. Users interact with the system by selecting the edge of the circular dial widget (4cm diameter) and dragging a cursor around its rim. The wheel color changes to indicate the direction of motion and as users move brief haptic or audio clicks are

played. The audio output is provided by standard earphones connected to the device's audio jack, while the tactile output is delivered via a matchbox sized SHAKE SK6 device capable of delivering a wide range of tactile cues [14]. The connection to the SK6 is achieved via a link to a PC (Wi-Fi) that communicates to the SHAKE device via Bluetooth. The SHAKE was manually mounted on the back of the phone with Velcro fasteners. The audio cues used in the system take the form of 113 ms audio *beeps* (Mono, 44100Hz, stored in a *wav* file). Analogously, the haptic cues are represented by sharp 50 ms vibro-tactile *buzzes*. These two cues were select to be short and distinctive via iterative, subjective testing by the authors during system development. Users are able to cancel a PIN entry at any time by shaking the device, a gesture captured from the built-in accelerometers.

The Spinlock GUI is composed of two screens, one to customize settings and the other to enter PINs. The first screen allows users to specify the length of the PIN (4-6 digits) and the direction-count pairs that compose it (numbers from 1 to 10 in either the clockwise or anti-clockwise direction). Connections to the host PC (via sockets for communication to the SHAKE device and data logging) are also managed on this screen. The PIN entry screen shows the input dial and a bar of colored rectangles which indicate PIN entry progress - grey for the number of PIN items entered, green for a correct complete PIN and red for a failed complete PIN.

4 Evaluation

Spinlock was evaluated with a user study. The goals were to compare performance between the two display modalities, to compare performance among PINs of varying complexity and to determine the resistance of the technique to observation attacks conducted via audio-visual recording equipment. Correspondingly, the study incorporated four conditions derived from two binary independent variables: modality and PIN complexity. The two modalities considered were *haptic* and *audio* cues, while the PIN complexity was manipulated by altering the data input range. This was achieved by varying the maximum number of cues that each PIN item could be composed of from five (*short:* each PIN item involved counting a number of cues in the range of between one and five inclusive) to ten (*long:* PIN items were from one to ten inclusive). Since each PIN item also includes an orthogonal binary direction component (left/right), the *short* PIN encompasses 10^4 possible combinations, equal to a standard 4 digit numerical PIN in terms of the level of security it provides against a brute-force (or PIN guessing) attack. The *long* PIN has 20^4 possible combinations, a significantly increased figure.

The study itself had a repeated measures design and involved 12 participants (seven male, five female with age between 22 and 30 years) each completing all four experimental conditions. PIN complexity was balanced among participants, with six completing each of the two possible orders. Modality was balanced within each PIN complexity block, such that three participants always started with haptics and three with audio. Each condition required participants make 15 successful PIN entries. The first five were considered practice and analysis restricted to subsequent interactions. Consequently data analysis took place on 40 correct PIN entries per user. As with most current ATM systems, each PIN was composed of four items so a total of 480

complete correct PIN entries and 1920 individual data inputs were examined. Erroneous input after completion of the practice trials was also analyzed.

The experiment was conducted in an empty office with participants seated in front of a desktop computer. After filling basic demographics and reading experimental instructions, they were shown the mobile device and provided with a randomly generated PIN in written form (either *short* or *long* depending on the order condition) and an experimenter demonstrated how to correctly enter a PIN. Participants were then given the opportunity to freely explore the system for a maximum of five minutes before the formal conditions commenced. All input took place on the mobile device, but experimental data was streamed to the desktop PC, which also displayed a window indicating the number of successful PIN entries required to complete the current condition. After completing one haptic and one audio condition, participants received a new randomly generated PIN (either *short* or *long* to complement their previous PIN) and used this for the remainder of the study. The experiment took 30-45 minutes in total.

Experimental measures were successful PIN entry time, error rate and the number of times users canceled a PIN entry process. Participants also completed a NASA TLX questionnaire directly after each condition. Data logging captured fine grained data relating to all user interactions. Finally, audio and video of the participants' hands and the mobile device running the experimental software were captured with a Sony camcorder mounted on a tripod and positioned directly over their shoulders.

5 Results

Experimental data are shown in Figure 3. All data were tested using two-way repeated measures ANOVAs. The authentication time performance attained a significant main effect of modality ($F(11, 1)=9.08$, $p=0.012$) and PIN complexity ($F(11, 1)=13.8$, $p=0.003$), but the interaction between these two variables was not significant ($F(121, 1)=0.28$, $p=0.6$). Authentication errors showed significant effect of modality

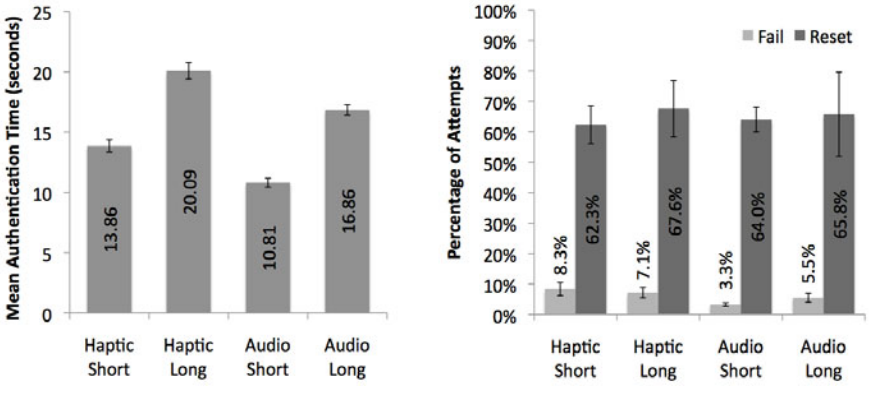

Fig. 3. Mean authentication time (left); mean percentage of failed trails and resets events (right). Bars show Std Error.

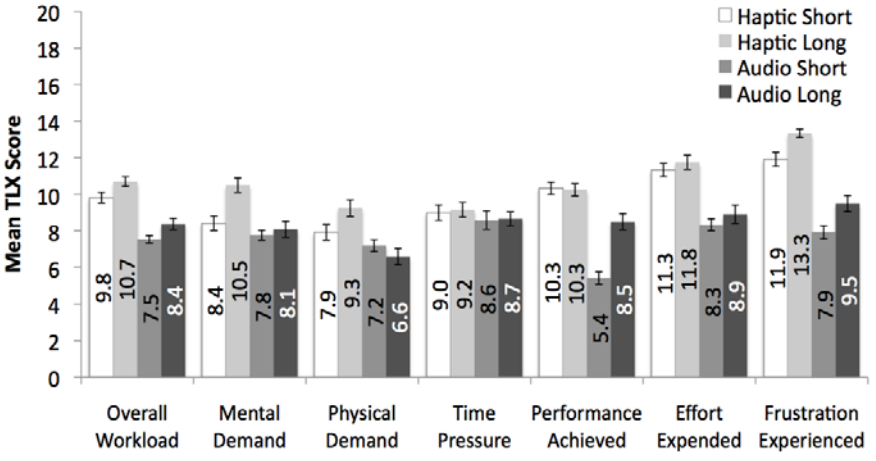

Fig. 4. NASA TLX ratings: higher scores show higher workload. Bars show Std Error

(F (11,1)=5.8, p=0.034) but not PIN complexity (F (11,1)=1.44, p=0.256) or interaction among the two (F (121,1)=0.66, p=0.433). Canceled PIN entries (resets) showed no significant variations across PIN complexity or modality (F (11,1)=2.65, p=0.13; F (11,1)=0.81, p=0.38). Finally, the two-way ANOVA on the overall workload of the TLX (Figure 4) showed a significant effect of modality (F (11,1)=15.23, p=0.002) but not PIN complexity (F (11,1)=3.7, p=0.081).

6 Discussion

This experimental work in this paper sought to explore how performance with the Spinlock system varied between haptic and audio cue presentation modalities and between PINs composed of more or less complex cues. The results clearly showed that participants found the haptic modality more challenging: significant differences were observed in the mean PIN entry times, failed authentication rates and overall workload. One possible explanation for this is system latency: the haptic effects were delivered on a wirelessly connected device while the audio cues were triggered *in-situ*. Although the impact of this cannot be determined by the current study, future work on this topic need more carefully control latency in the display of haptic cues.

PIN complexity, on the other hand, resulted in increased task completion times, but had no significant effect on other metrics. The increased time is unsurprising in this case: compared to the *short* complexity condition, participants had to make larger input strokes in the *long* complexity condition. The fact that the increase in complexity did not result in increases in the error rate or levels of workload strongly suggests that the task of counting the haptic and audio cues is easy to understand, effective and scalable. This is an encouraging result.

Analyzing the erroneous PIN entry trials also provided valuable insights into participant performance. In these trials, no errors of direction of travel were made and 82% of error trials involved a mistake in only one PIN item (from the four composing

each PIN). Also, the majority of errors (78%) involved entering digits one higher or lower than the target item. Comments by participants provided a feasible explanation for this; several spontaneously remarked that the randomly distributed nature of the cues made predicting the location of the final target challenging. In particular, several mentioned that unintentionally overshooting the target item was the most frustrating aspect of the experiment. That participants were typically aware of such errors, rather than unaware, is evidenced by the relatively high number of manual reset events - participants realized they had erred and immediately cancelled the trial. Participants also proposed strategies for mitigating this effect, including increasing the minimum spacing between cues, randomizing cue spacing per PIN item rather than per cue, accepting one item beyond the target as valid input (e.g. if the target is 4-left, accept both 4-left and 5-left as valid) and providing a mechanism for re-entering a single PIN item. Several participants also commented that although they felt the audio interface was "easier", they preferred the haptic version as it was more "private".

Spinlock also performs well compared to previous systems reported in the literature. For example, PhoneLock [9], an authentication system based on the recognition of a set of tactile or audio cues achieves mean authentication times and error rates of 18.7 seconds and 7% compared to the 15.4 seconds and 6% observed in the current study. Considering haptic performance alone, mean task time in the Spinlock system improves 30% over that reported in PhoneLock (16.9 seconds vs. 24.05 seconds). These results suggest that systems that rely on counting haptic cues may be more effective that those that rely on tactons, at least in some scenarios.

7 Security Analysis

By relying on the perception of non-visual cues Spinlock obfuscates its data input process - unlike keypad systems for PIN entry, simply looking at a user's hands whilst they are entering data does reveal the PIN contents. The randomization of spacing between the cues delivered by the system was intended to reinforce this and reduce the relationship between the user's observable input and the PIN item they enter. However, an analysis correlating PIN item entry time with PIN item number across all four experimental conditions was significant ($r(28) = 0.87$, $p<0.001$) indicating this manipulation was not fully successful and representing a security threat.

To gauge the severity of this threat, an expert with full knowledge of the Spinlock system performed an observation attack using the complete set of experimental videos taken from two randomly chosen participants (a total of 80 authentications using four PINS and both modality and complexity conditions). To facilitate the attack, the expert was provided with a table indicating mean selection times for each PIN item. The expert was unable to correctly deduce any of the four PINs studied and reported that determining a PIN from a single observation would be impossible. However, the repeated presentation of each PIN 20 times enabled trends to be ascertained. In particular, the expert performed well with PIN items with low digits (rapid trials) and was able to easily isolate (although not precisely ascertain) input relating to high digits. Audio cues from the camera's microphone were not reported to contain any useful information - even in the stable, quiet lab environment the attacker stated that

neither the audio cues to the headphones or the vibrations to the SHAKE produced any environmentally audible noise.

8 Conclusions and Future Work

The contribution of this work is the presentation of a novel design for a haptic and audio PIN entry system. It combines the simplicity of previous approaches based on simply recognizing cues (rather than applying further mental mappings or transformations to the perceived information [e.g. 4, 6]) but avoids the overhead of requiring users to learn and recognize a large stimuli set [e.g. as in 3, 8, 9]. It achieves this by asking users to count, rather than accurately distinguish, the number of simple, identical haptic or audio cues that are delivered in response to their input.

A prototype instantiating this idea, Spinlock, was developed and a preliminary evaluation performed. The results show this approach has considerable promise and suggest it can reduce the high levels of cognitive load (and associated task times and error rates) observed in studies of previous non-visual PIN entry systems. The study also suggests fruitful avenues for future work, including potential revisions to the interaction design that provide better error prevention or recovery mechanisms and the need to address a security weakness to repeated observations through the development of improved randomization functions for stimuli presentation. Further empirical user studies and security analyses to validate such refinements are also required.

Acknowledgments. Partial support for this research was provided by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through the Carnegie Mellon | Portugal Program. We would also like to thank our experimental participants.

References

1. Giesen, L.: ATM fraud: Does it warrant the expense to fight it? Banking Strategies 82(6) (2006)
2. De Luca, A., Langheinrich, M., Hussmann, H.: Towards understanding ATM security: a field study of real world ATM use. In: Proceedings SOUPS 2010 (2010)
3. Bianchi, A., Oakley, I., Kwon, D.S.: The Secure Haptic Keypad: Design and Evaluation of a Tactile Password System. In: CHI 2010, pp. 1089–1092. ACM, New York (2010)
4. Sasamoto, H., Christin, N., Hayashi, E.: Undercover: authentication usable in front of prying eyes. In: Procs of CHI 2008, pp. 183–192. ACM, New York (2008)
5. Brewster, S.A., Brown, L.M.: Non-visual information display using tactons. In: Procs of CHI 2004 Extended Abstracts, pp. 787–788 (2004)
6. Malek, B., Orozco, M., Saddik, A.: Novel shoulder- surfing resistant haptic-based graphical password. In: Proceedings of EuroHaptics (2006)
7. De Luca, A., von Zezschwitz, E., Hußmann, H.: Vibrapass: secure authentication based on shared lies. In: Procs. of CHI 2009, pp. 913–916. ACM, New York (2009)

8. Bianchi, A., Oakley, I., Lee, J., Kwon, D.: The haptic wheel: design & evaluation of a tactile password system. In: Proceedings of CHI 2010, pp. 3625–3630. ACM, New York (2010)
9. Bianchi, A., Oakley, I., Kostakos, V., Kwon, D.: The Phone Lock: Audio and Haptic shoulder-surfing resistant PIN entry methods. In: Proc. of ACM TEI 2011. ACM, New York (2011)
10. Kuber, R., Yu, W.: Feasibility study of tactile-based authentication. International Journal of Human-Computer Studies 68(3), 158–181 (2010)
11. Brown, L.M., Brewster, S.A., Purchase, H.C.: Purchase, Multidimensional tactons for non-visual information presentation in mobile devices. In: Proc. of MobileHCI 2006, pp. 231–238 (2006)
12. Garcia-Salicetti, S., Beumier, C., Chollet, G., Dorizzi, B., Jardins, J., Lunter, J., Ni, Y., Petrovska-Delacrétaz, D.: BIOMET: A Multimodal Person Authentication Database Including Face, Voice, Fingerprint, Hand and Signature Modalities. In: Kittler, J., Nixon, M.S. (eds.) AVBPA 2003. LNCS, vol. 2688, pp. 845–853. Springer, Heidelberg (2003)
13. Faraj, M.I., Bigun, J.: Audio-visual person authentication using lip-motion from orientation maps. Pattern Recognition Letters 28(11), 1368–1382 (2007)
14. SHAKE SK6, `http://code.google.com/p/shake-drivers`