Interacting with Computers xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Interacting with Computers

journal homepage: www.elsevier.com/locate/intcom

Counting clicks and beeps: Exploring numerosity based haptic and audio PIN entry

Andrea Bianchi [a,*], Ian Oakley [b], Dong Soo Kwon [a]

[a] KAIST, 291 Daehak-ro (373-1 Guseong-dong), Yuseong-gu, Daejeon 305-701, Republic of Korea
[b] M-ITI, University of Madeira, Campus da Penteada, 9020-105 Funchal, Portugal

ARTICLE INFO

Article history:
Received 20 October 2011
Received in revised form 10 June 2012
Accepted 12 June 2012
Available online xxxx

Keywords:
Human factors
Haptics and audio
Authentication
Mobile

ABSTRACT

Haptic and audio cues now appear commonly in computer interfaces, partially due to inherent advantages such as their support for eyes-free interaction. Their invisible, unobservable nature also makes them ideal candidates for security interfaces in which users have to enter secret information such as passwords. In particular, researchers have explored this idea through the design of PIN entry authentication systems based on multi-modal combinations of visual and non-visual content or on the recognition of small sets of unimodal haptic or audio stimuli. This paper highlights the benefits and performance limitations of these approaches and introduces an alternative based on unimodal audio or haptic temporal numerosity – the ability to accurately and rapidly determine the number of cues presented in rapid temporal succession. In essence, in a numerosity interface, rather than recognizing distinct cues, users must count the number of times that a single cue occurs. In an iterative process of design and evaluation, three prototypes implementing this concept are presented and studies of their use reported. The results show the fastest PIN entry times and lowest error rates to be 8 s and 2%, figures that improve substantially on previous research. These results are attained while maintaining low levels of workload and substantial resistance to observation attack (as determined via camera attack security studies). In sum, this paper argues that unimodal audio and haptic numerosity is a valuable and relatively unexplored metaphor for non-visual input and demonstrates the validity of this claim in the demanding task of unobservable authentication systems.

1. Introduction

Modern computer interfaces place a strong emphasis on information conveyed in the visual modality, a fact aptly illustrated by currently dominant and highly successful paradigms such as Graphical User Interfaces (GUIs). Although there is a longstanding research interest in investigating the feasibility of non-visual modalities to support novel user-centered interaction techniques (e.g. Kortum, 2008), most researchers still approach sensory channels such as audio and touch (or haptics) from the perspective that they are auxiliary modalities capable of supporting or facilitating primarily visual tasks (Oviatt, 1999). Perhaps the most common use of such non-visual cues is as re-enforcement of visual messages through simultaneous redundant presentation: consistent, non-interfering messages delivered to two or more senses. Supported by well developed constructs such as Multiple Resource Theory (Wickens, 2008), this approach has been explored in domains as diverse as medical training (Boulanger et al., 2006), rehabilitation systems (Banala et al., 2008), immersive virtual environments (Reiner, 2004), driving simulators (Ho et al., 2005; Bernhard et al., 2009; Chang et al., 2008) and performative musical instruments (Gallace et al., 2007). Typical motivations for such work include to create more "natural" interactions or an explicit argument that redundant information presented to multiple sensory channels will lead to faster task execution times or lower error rates (Pirhonen and Tuuri, 2008).

Recently, the use of audio and, in particular, haptic cues have been explored in the domain of security and specifically in the design of authentication interfaces for public spaces, such as Personal Identification Number (PIN) entry processes on bank Automatic Teller Machines (ATMs) or passcode door locks (e.g. Sasamoto et al., 2008; Bianchi et al., 2011b). A key motivator for this work is the fact that audio and haptic cues are inherently secure against visual observation; unlike standard input systems involving pressing buttons or keys, they rely on information that cannot be seen. Taking advantage of this fact, researchers have designed a range of multimodal systems (e.g. Sasamoto et al., 2008; De Luca et al., 2009) that use non-visual cues to obfuscate PIN entry processes at public terminals with the objective of defeating observation attacks (De Luca et al., 2009) – malicious attempts to obtain users' PINs either by surreptitious in-person surveillance (shoulder surfing) or via recordings made by appropriately positioned equipment (camera attacks). However, the non-visual information deployed in

* Corresponding author. Address: Human Robot Interaction Center, Dept. of Mechanical Engineering, KAIST, 355 Gwahangno, Yuseong-gu, Daejeon 307-701, Republic of Korea. Tel.: +82 010 2317 5220.
E-mail addresses: andrea.whites@gmail.com, andrea@kaist.ac.kr (A. Bianchi).

0953-5438/$ - see front matter © 2012 British Informatics Society Limited. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.intcom.2012.06.005

such multimodal systems represents, by definition, a unique component of the interface rather than a redundant presentation of visual cues. Consequently, such systems have proven challenging for users, and have led to PIN entry times in excess of 30 s (Sasamoto et al., 2008). Attempting to address these performance issues, researchers have also explored the design of unimodal non-visual PIN entry interfaces. Evaluations have shown these provide substantially improved performance (e.g. Bianchi et al., 2010a, 2010b). An explanation for the superiority of unimodal solutions are findings in cognitive science that suggest that users engaged in cognitive task with high demands on attention perform better when they are not required to split their focus over multiple sensory channels (Spence and Driver, 1997).

Unimodal audio and haptic authentication systems (e.g. Bianchi et al., 2010a; Kuber and Yu, 2010) are generally based on the recognition of structured tactile or audio cues, known respectively as tactons (Brewster and Brown, 2004) and earcons (Brewster et al., 1993). PINs are constructed as a sequence of such cues, in much the same way that standard ATM PINs are constructed from a sequence of numbers. Consequently, in order to enter such a PIN, users need to perform a range of cognitive and perceptual tasks including learning an alphabet of non-visual cues, composing and recalling a PIN based on a sequence of those cues and rapidly and reliably recognizing and identifying each cue. This paper argues these tasks are challenging and still poorly understood and its contribution lies in proposing an alternative approach to unimodal authentication that makes fewer requirements on users' mental resources. This approach is inspired by haptic and audio temporal numerosity (Gallace et al., 2007) – the ability to accurately, confidently and rapidly determine the number of cues presented in rapid temporal succession. The paper includes a practical contribution via the design and development of Colorlock and Timelock, two novel prototypes demonstrating, exploring and refining non-visual PIN entry processes based on numerosity. Finally, it includes a user-centric contribution achieved through empirical user evaluations of simulated PIN entry processes in Colorlock and Timelock that demonstrate the viability and security of the counting approach and discussions that contrast it to the state of the art.

The remainder of this paper is structured as follows: a literature review covering both recognition and numerosity in non-visual modalities and a discussion of haptic and audio based authentication systems; a description of the abstract interaction model underlying the cue-counting authentication technique, including an in-depth summary of SpinLock (Bianchi et al., 2011b), a recent example of a system deploying such an approach; the description and evaluation of Colorlock and Timelock, two novel systems. The paper concludes with a general discussion of cue-counting haptic and audio PIN entry systems and the wider applicability and prospects of this style of non-visual interaction.

2. Related work

2.1. Haptic and audio recognition and numerosity

Researchers have long sought to optimally convey information through non-visual channels. Two modalities have been widely considered: audio, delivered via standard speakers or headphones and touch, delivered via a wide range of actuator technologies. The community studying touch interaction follows conventions in experimental psychology and uses the term "haptics" to refer to a wide range of tactile and kinesthetic experiences and to the research addressing them (Robles-De-La-Torre, 2006; Henriques and Soechting, 2005). In the past few decades, this term has been extended to apply to sensing and display systems, leading to terms such as *haptic interfaces*, *haptic rendering* and *haptic applications* (Biggs and Srinivasan, 2002). This paper adopts this widespread terminology.

Broadly speaking, two strategies exist for information display via the audio and haptic modalities. The first is based on recognition of complex iconic or symbolic cues and has received considerable attention. The second is based on numerosity (or ability to reliably count or ascertain the quantity) of simple identical cues. This approach has been relatively overlooked. Literature relating to both techniques is reviewed below.

Perhaps the most commonly studied way to embed information within haptic and audio stimuli is by iconically associating cues with concepts – by ensuring that audio or haptic stimuli used are evocative and reminiscent of the concepts. For example, the sound of crushed paper has been associated with a delete operation (Gaver, 1989) and increased friction proposed as a technique to indicate larger file sizes (Bau et al., 2010). These types of mapping are referred to as auditory icons (Gaver, 1989), defined as auditory representation of objects or notions that embody a literal, intuitive meaning, or haptic icons (MacLean, 2003), defined as brief computer generated signals displayed to a user through force or tactile feedback to convey information such as event notification, identity, content or state. These are powerful techniques that have proven highly successful in the design of audio cues, many of which are now integrated into modern computer interfaces. Work on haptic icons has achieved less practical success, partly due to the specialized hardware required to render such cues and partly due to the challenge in creating meaningful, iconic haptic sensations for a wide range of operations. Although guidelines exist for the design of haptic icons (van Erp, 2002), it remains a challenging and subjective process (van Erp, 2002; Ternes and Maclean, 2008).

Knowing that iconic cue design is not always achievable, researchers have also explored symbolic cue designs in both audio and haptics. Respectively termed earcons (Brewster et al., 1993) and tactons (Brewster and Brown, 2004), this has involved the creation of stimuli whose meaning is not naturally known to users but must be learned. Earcons are constructed using combination of pitch, timbre, register, rhythm and intensity (Brewster et al., 1993), while tactons have been built with properties such as roughness, rhythm and amplitude (Brown et al., 2005; Qian et al., 2009). Studies have suggested that such structured cues are cross-modal in nature, that equivalent pairs of earcons and tactons can be effectively identified and recognized (Hoggan and Brewster, 2007). Experiments have shown recognition rates for earcons and tactons of approximately 70% (Brewster et al., 1993; Brown et al., 2005) with an alphabet of five or less stimuli. Although this is sufficient for a wide range of application scenarios, it is arguably insufficient for authentication processes, where a PIN typically consists of a unique sequence of items from a set of 10,000 or more possible sequences.

Counting the number of short, pulse-like stimuli in a sequence is an alternative mechanism for communicating structured non-visual information. A common example of this technique is found in safe lock dials, which require users to count the number of ticks they move beyond a pivot point, or in the clicks delivered to delimit and separate menu items in haptics-enabled rotational dials such as the BMW iDrive (Bernhard et al., 2009). The ability these systems tap into is the numerosity (Gallace et al., 2007): the human ability to accurately sense and ascertain the number of rapidly sequential or simultaneous (but spatially distributed) cues. High levels of numerosity performance have been observed over the senses of vision, audio and touch, leading researchers to conclude that numerosity is amodal and largely dependent on cognitive process above the perceptual level (White and Cheatham, 1959). However, research has also suggested that people perform better at uni-modal rather than multi-modal numerosity tasks, perhaps due to incompatibilities in crossmodal integration or sharing of

Please cite this article in press as: Bianchi, A., et al. Counting clicks and beeps: Exploring numerosity based haptic and audio PIN entry. Interact. Comput. (2012), http://dx.doi.org/10.1016/j.intcom.2012.06.005

cognitive/attentional resources (Spence et al., 2001). In an explicit comparison of numerosity performance over different modalities, (Lechelt, 1975) reported that the most significant errors were a tendency to over-estimate the number of cues presented, something that was more extreme in visual tasks than in audio or haptic ones. However, in practice, with inter-cue intervals of 320 ms or more, Philippi et al. (2008) reports very low error rates across modalities for up to nine items, and that up to five haptic and audio cues can be counted with high accuracy providing the inter-cue interval is 160 ms or greater. Finally, it has been reported that subitising (rapid, accurate detection of the number of stimuli in groups of five or less) is possible for the audio modality (ten Hoopen and Vos, 1979) but not for haptics (Gallace et al., 2008). This paper argues that the high levels of performance seen in non-visual numerosity tasks, and the relative lack of literature considering this task in the design of interactive systems, makes it a novel and suitable candidate for exploration in the domain of non-visual authentication and security interfaces.

2.2. Haptic and audio authentication

Researchers have studied a wide range of haptic and audio PIN entry techniques. The early work in this field adopted a multimodal approach, combining the rich visual modality of graphical or alphanumerical passwords with haptic information. For instance, Malek et al. (2006) described haptic passwords that used pressure-based input as hidden channel to obfuscate a graphical input. In this system users operated a stylus on a touch sensitive display. They drew a graphical password composed of lines connecting points on a grid while systematically applying different levels of pressure to the different points. The final result was an easy-to-remember graphical password augmented by an hidden channel of haptic pressure information, a technique that makes observation highly challenging.

Attempting to adapt this multi-modal approach to more standard input techniques based on button presses at public terminals, Sasamoto et al. (2008) and De Luca et al. (2009) respectively presented Undercover and Vibrapass, two PIN entry interfaces that use a combination of observable visual input and unobservable haptic cues as a hidden channel. Undercover is based on a graphical password system that requires the users to select a sequence of pre-determined images, but that obscures the mapping between the interface buttons and the possible answers via directional information rendered on a haptic device hidden by the user's hand. Vibrapass obfuscates a visually observable numeric PIN entry process by asking users to input a mix of correct information and lies (intentionally incorrect input). The system instructs the user when to enter correct or incorrect information via tactile cues delivered to a personal mobile device that has been pre-paired with the terminal. Although promising and effective, these approaches require users to invest significant cognitive resources in order to sense and recognize the haptic stimuli and then map them to the correct actions. 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 s, with error rates of between ~26–52% (Sasamoto et al., 2008).

In contrast to this multi-modal approach Bianchi et al. (2010a) proposed a uni-modal haptic password based on recognizing and selecting a sequence of tactons on a special keypad capable of rendering haptic stimuli. To secure against observation, the tactons were randomized over the keys after every input from the user. As the users' task consists of searching for, recognizing and selecting haptic cues, Bianchi suggests it will induce lower levels of cognitive load (corresponding to improved task completion times and error rates) when compared to multi-modal approaches. Subsequent evaluations of a number of system variations (Bianchi et al., 2010b), including equivalently constructed audio systems (Bianchi et al., 2011a), support this claim and show authentication times of approximately ~12–19 s with error rates of ~5–7% (Bianchi et al., 2011a). These claims are supported by closely related work by Kuber and Yu (2010) that explores the use of spatially distinct cues rendered on Braille cells for accessible password entry tasks and achieves similar findings. One disadvantage of these unimodal systems is that they are reliant on users accurately selecting specific haptic cues from a set of possible stimuli, a challenging task when sets exceed 3 or 4 items in size (e.g. Bianchi et al., 2010b, Brown et al., 2005). Indeed, more generally, issues of recognizing, learning and memorizing tactons all remain relatively poorly understood, and it is currently unclear if purely haptic PINs are a reliable, scalable and feasible concept (Brown et al., 2005). These issues place doubts on the viability of recognition-based approaches to haptic authentication.

In summary, this review has covered general techniques for presenting information in haptic and audio modalities and explored specific examples in which these modalities have been deployed in authentication interfaces. It concludes that counting or numerosity is a promising paradigm for non-visual information presentation and one that is both suitable for, and relatively unexplored in, the domain of haptic and audio authentication interfaces. The remainder of this paper serves to address this topic and omission.

3. Haptic and audio single-cue PIN entry

This paper proposes a novel technique for non-visual PIN entry based on numerosity, or counting. It argues that, compared to recognition based approaches, numerosity tasks simplify the perceptual and cognitive work users must complete in order to successfully enter a PIN; Abstractly, the numerosity interaction technique operates as follows. The user first initiates the display of cues via an explicit command. A sequence of cues is then presented to the user in a rapid, but randomized temporal distribution. The randomization of cues serves to obscure the data-entry process from observers. When the user has sensed (and counted) the number of cues required to enter their current PIN item, they issue an explicit command to mark this point and halt further display of cues. The cue-count is then recorded by the system.

In this paper we describe specific interfaces implementing this technique: SpinLock, Colorlock and Timelock. SpinLock is presented as a detailed summary of prior, highly related work (Bianchi et al., 2011b) and is included for completeness. It is based on the delivery of non-visual cues in response to the spatial distance users travel while executing a continuous circular gesture. In contrast, Colorlock and Timelock are novel prototypes that deliver non-visual cues in response to dwell-time on GUI input buttons. The three prototypes also systematically vary in the randomization procedures used to obfuscate the cue presentation process and in the way they combine the non-visual cues with visually observable delimiting input (e.g. the direction of the circular gesture or the particular buttons selected). The Timelock prototype also introduces a UI metaphor that supports a range of error correction and recovery techniques and evaluates how these are employed by users. Taken together these three systems provide a thorough exploration of the design space of using non-visual cue counting for PIN entry tasks. Evaluations of these systems reveal how this approach compares with prior systems in terms of task completion time, error rate and subjective satisfaction and workload.

Please cite this article in press as: Bianchi, A., et al. Counting clicks and beeps: Exploring numerosity based haptic and audio PIN entry. Interact. Comput. (2012), http://dx.doi.org/10.1016/j.intcom.2012.06.005

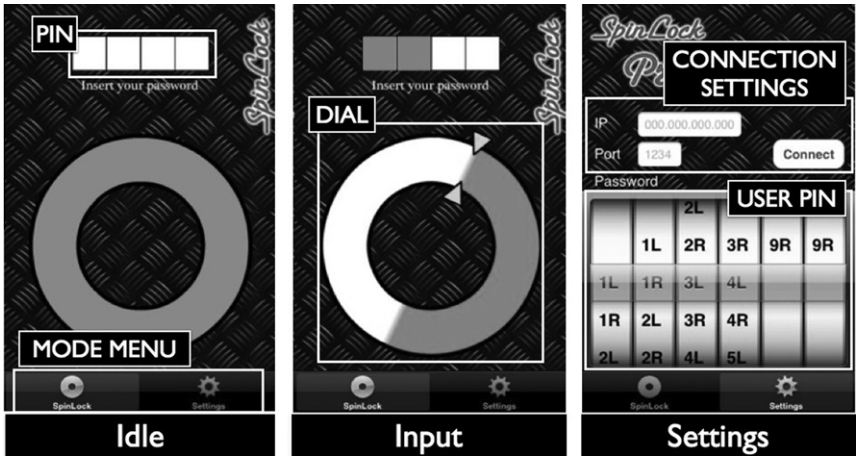

Fig. 1. The Spinlock Graphical User Interface: whilst idle (left), during the user interaction with two PIN items entered (center) and the settings screen showing user password (right).

4. Spinlock: design and evaluation

SpinLock (Bianchi et al., 2011b) is based on the rotary dial-lock of a traditional safe. Such systems are unlocked by inputing a sequence of numbers statically printed on a dial in alternating clockwise and anti-clockwise directions. An example safe PIN would be 2-anti-clockwise, 8-clockwise, 5-anti-clockwise, and 7-clockwise. Spinlock borrows this interaction scheme, but removes the requirement of alternating the direction of motion between individual entries. It also replaces statically marked numbers with the action of counting audio or haptic cues delivered during the rotation interaction. In order to remain resistant to observation the spatial distance users must travel between cue presentations is randomized after every cue. Consequently, the distance that dial is rotated does not directly correspond to the counting data that is input.

Spinlock was implemented for the Apple iPhone and iPod Touch devices (Fig. 1). The touch screen was used for input. Users interact with the system by selecting the edge of the circular dial widget (4 cm diameter) and dragging a cursor around its rim. The audio output was provided by standard earphones connected to the device's audio jack, while the tactile output was delivered via a matchbox sized SHAKE SK6 device capable of delivering a wide range of tactile cues and attached on the back of the device. The connection to the phone is achieved via a link to a PC (Wi-Fi) that communicates to the SHAKE device via Bluetooth.

Spinlock was evaluated (Bianchi et al., 2011b) with a user study (summary in Table 1). The goals were to compare performance between two display modalities (haptic vs. audio), to compare performance among PINs of varying complexity (a simple PIN consisting of numbers ranging from 1 to 5, and a complex PIN with numbers from 1 to 10) and to determine the resistance of the technique to observation attacks (via camera attack over multiple PIN entries). The study was composed of four conditions, used a repeated measures design and a Latin square approach to balance order and practice effects. During an initial session, 12 participants were assigned randomly generated PINs, memorized them and practiced with the system. Directly afterwards they completed 40 correct PIN entries (10 per condition). 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. Data from a NASA TLX (Hart and Staveland, 1988) was also collected. Video of the study was recorded in order to perform a camera attack.

Overall data showed good task completion times and error rates (Figs. 2 and 3): overall means of 12.3 s and 5.8% (simple PIN), and 18.4 s and 6.3% (complex PIN). Results also indicated participants found the haptic modality more challenging than the audio one: significant differences were observed in the mean PIN entry times, failed authentication rates and overall workload (refer to Bianchi et al. (2011b) for details). A possible cause of this is latency in the Bluetooth communication used during display of the haptic cues. PIN complexity resulted in increased task completion times, but had no impact on error rate and resets suggesting the counting task is scalable between 5 and 10 items. In terms of errors, no error was caused by incorrectly selecting the direction of motion, and 82% of error trials involved a mistake in only one of the four PIN items. Furthermore, 78% of the errors involved entering a single digit higher or lower than the target item. In interviews, participants indicated that overshooting the target item was the most frustrating aspect of the experiment and suggested strategies for mitigating this effect, including increasing the minimum spacing between cues, randomizing cue spacing per PIN rather than per PIN item, accepting one item beyond the target as valid input, and providing mechanisms for re-entering a single PIN item.

A limited camera attack executed by a security expert was conducted based on 80 PIN entries from two users. It covered all four experimental conditions and four associated passwords. The results revealed that although SpinLock was much more resistant to observation that standard PIN input, a dedicated attacker would be able to uncover the PIN given a sufficient number of observations – no PINs were uncovered, but trends in the process were clear. This suggestion was confirmed by the fact that the correlation between PIN item entry time and PIN item number was significant ($r(28) = 0.87$, $p < 0.001$). Together, these results indicate that the SpinLock design is resistant to one-off observation, but susceptible to repeated observation attacks.

5. Prototype I: Colorlock

5.1. Design and implementation

The design of Colorlock continued exploring PIN entry via counting rather than recognizing non-visual cues and instantiated lessons learned from the SpinLock study. Once again, both haptic and audio cues were considered. However, in Colorlock, these cues were triggered not by movement, but by *dwell time* – users pressed

Please cite this article in press as: Bianchi, A., et al. Counting clicks and beeps: Exploring numerosity based haptic and audio PIN entry. Interact. Comput. (2012), http://dx.doi.org/10.1016/j.intcom.2012.06.005

Table 1
A summary of the experimental settings, conditions and methods used for testing the Spinlock, Colorlock and Timelock prototypes.

Interface name	Spin lock	ColorLock	TimeLock
Number of random PINs given to users (required to be memorized)	2 (one for each of the the secondary conditions)	1	1
PIN length	4		
Security (brute force)	1 in 10,000 or less		
Security (observation)	1 in 10.000 or less	1 in 625	1 in 625
Modality condition	2 (haptic vs. audio)		
Secondary condition	2 (simple PIN vs. complex PIN)	2 (constant vs. random beat)	2 (constant vs. random beat)
Number of participants	12		
Training trials per session (users can look at the PIN)	5		
Experiment trials per session (users must have memorized the PIN)	10		
Total number of trials per user	40 (4 conditions × 10 experiment trials)		
Total number of sessions for analysis	480 (40 sessions × 12 users)		
Notes	• Partially balanced study (Latin-square design) •PIN were shown at the beginning of a session for being memorized • Post hoc interview and TLX		

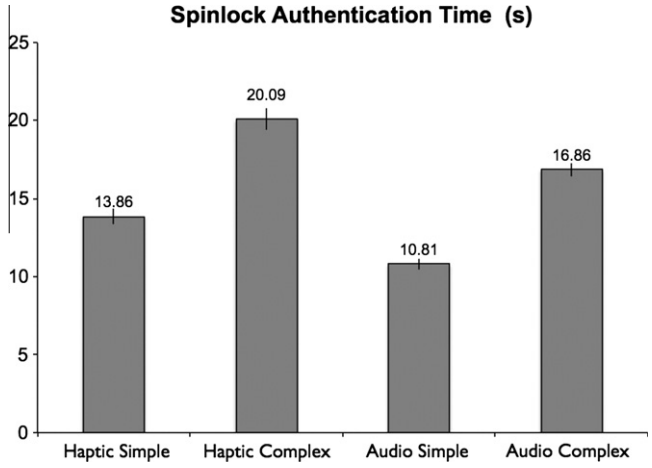

Fig. 2. Spinlock authentication time (in seconds).

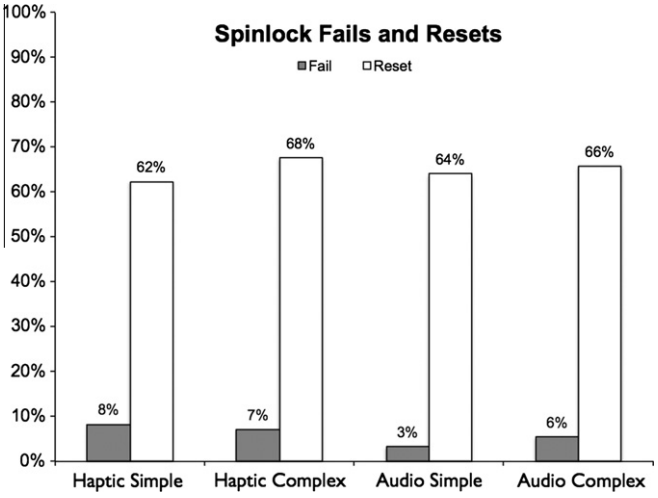

Fig. 3. Spinlock fails and resets.

and held an on-screen button for a prolonged period (~2–4 s) and sequential cues were triggered at randomly determined intervals during this dwell action. The Colorlock user interface contained a total of four buttons with this behavior, each a large, distinctively colored target that occupied a fixed quadrant of the screen, as shown in Fig. 4. The four colors[1] used were: red, blue, green and yellow.

The four buttons were used to increase the entropy of input in the system. PINs in Colorlock were composed of a combination of visually observable information, in the form of selecting one of the colored buttons, and non-visual information, in the form of counting haptic or audio cues. Hence, a Colorlock PIN was composed of color-number pairs such as red-5, green-2 or blue-4. A minimum of one and a maximum of five non-visual cues were used to compose PIN items. This orthogonal combination of four possible colors and five possible numbers, led to 20 possibilities for each PIN item. Consequently, over a four item PIN, a total entropy of 160,000 (20^4) PIN combinations was achieved against bruce-force attack; a smaller entropy of 625 (5^4) was achieved against observation.

Cue randomization in Colorlock was achieved via two methods: *constant beats* or *random beats*. For constant beats, the inter-cue interval was randomly selected once per PIN item – the spacing between successive beats was identical. On the other hand, for random beats the interval was re-computed every time a cue was triggered. In both methods, the time from button press to the first beat was randomly set between 0 and 1200 ms (a subjectively chosen intervals intended to minimize lengthy initial pauses) while the inter-cue interval was limited to between 300 (the lower limit for perfect haptic and audio numerosity (Philippi et al., 2008) and 600 ms.

Colorlock was developed for Google's Android operating system and specifically on the Samsung Galaxy Tab. This device was selected due to its wide range of input and output features. These include an inbuilt vibrotactile actuator and APIs to control it, a headphone jack for audio display and a large 7 in. touch screen well suited to the presentation of the four colored buttons that form the basic Colorlock UI. Most importantly, this choice of the device ensured low latency delivery of haptic cues, addressing a potential problem identified with the SpinLock system. The audio cue used in the system took the form of a standard beep and vibration buzz as previously used in similar research (Bianchi et al., 2011a,b). The audio beep is 113 ms long, sampled as Mono 44100 Hz, played as 35–50 dB in earphones, with major frequency components of f1(F6) f2(c8) f3(C9#) f4(D9), and stored in a wav file. The haptic cue is a 25 ms vibration buzz produced through the activation of the standard built-in linear vibration motor (empirical measures: 1.25–1.8G, response time 0.025 ms, resonant frequency 170–250 Hz, noise 50 dB or less). The selection of these specific cue parameters was achieved via an iterative subjective

[1] For interpretation of color in Fig. 4, the reader is referred to the web version of this article.

Please cite this article in press as: Bianchi, A., et al. Counting clicks and beeps: Exploring numerosity based haptic and audio PIN entry. Interact. Comput. (2012), http://dx.doi.org/10.1016/j.intcom.2012.06.005

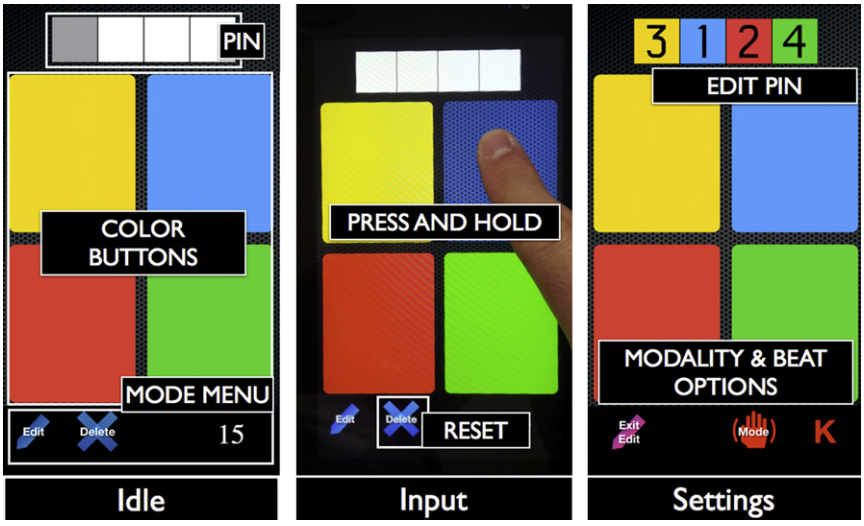

Fig. 4. The Colorlock Graphical User Interface: whilst idle (left), during the user interaction (center) and the settings screen showing user password (right).

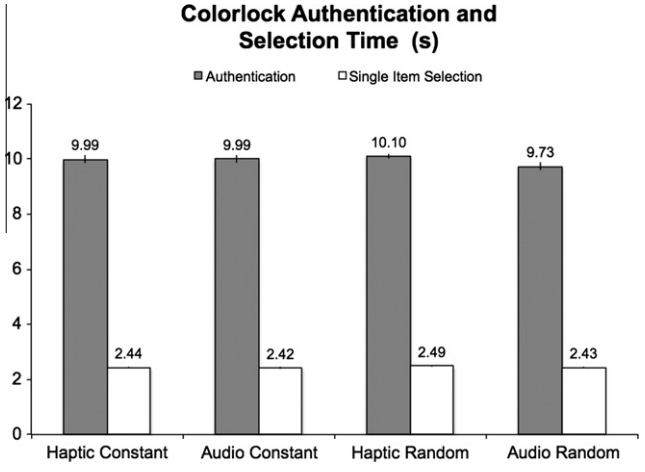

Fig. 5. Colorlock authentication and PIN item selection time (seconds).

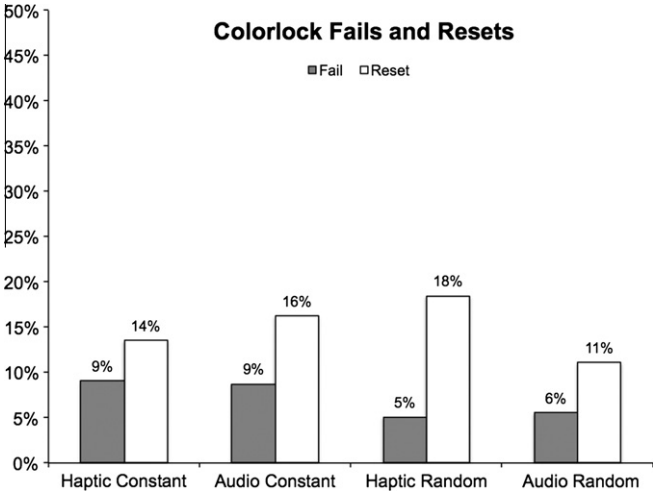

Fig. 6. Colorlock fails and resets.

process of trial and error with the target hardware, in order to minimize duration and magnitude while ensuring perceptibility.

Practically, the Colorlock graphical interface was composed of two screens: the application screen and the setting screen (Fig. 4). In the application screen the majority of the display shows the four colored buttons described above. At the top of the screen, four color-encoded boxes represent the PIN status with their color: white signifies an unentered item, grey an entered item, green a successful complete PIN and red an incorrect complete PIN. At the base of the screen are buttons to erase the current entered PIN items, to access the settings screen and a display of the number of PIN items entered. The settings screen shows a menu to choose modalities (haptic or audio) and beats (constant or random) and a widget to edit the target PIN.

5.2. Evaluation

Colorlock was evaluated with a user study. The goals were to compare performance between the two display modalities (haptic and audio) and two beating schemes (constant and random).

5.2.1. Participants

Twelve participants (eight male, four female with age between 20 and 36 years, mean: 26.4 and SD: 4.9) completed the study. They were a mix of researchers, students and professionals. Thirty-three percent reported to be advanced computer users, 50% medium and 17% basic. None of the participants were involved in previous related experiments. Participants were compensated for their time with a small gift.

5.2.2. Materials and methods

This study (summary in Table 1) compared four conditions derived from two binary independent variables: modality (haptic/audio) and beating type (constant/random). The study had a repeated measures design balanced according to a Latin square scheme. Modality was completely balanced among participants and beating type balanced within each modality block. Each condition required participants to make 15 successful PIN entries, the first 5 of which were considered as training trials and not analyzed. Consequently, a total of 40 correct PIN entries were collected per user, for a total of 480 complete correct PIN entries and 1920 individual PIN items.

The experiment was conducted in a quiet room. After filling basic demographics and reading experimental instructions, participants were shown the mobile device and provided with a randomly generated PIN. An experimenter demonstrated the system and participants had the chance to familiarize themselves

with its operation for a maximum of 5 min. Participants were required to memorize the PIN during this exploratory session and the training trials. The experiment commenced immediately afterwards and took on average 25–30 min to complete. All input took place on the Galaxy Tab device, where the user input was also recorded and logged in files.

Experimental measures were PIN entry times for successful authentication, error rate and the number of times users canceled a PIN entry process (subsequently called *resets*). Participants also completed a NASA TLX (Hart and Staveland, 1988) questionnaire directly after each of the four conditions. Moreover, videos showing the hands and device screen in close-up were recorded for all users for the duration of the study. Finally, the study closed with a short interview to collect preferences and feedback from users. Although general comments were solicited, the interview always included the following specific questions: which modality (haptics vs. audio) do you prefer?; which beating scheme (random vs. constant) do you prefer?; and was the PIN easy to memorize?

5.3. Results

Experimental data are shown in Figs. 5 and 6. All data were tested using two-way repeated measures ANOVAs. There were no significant differences in the time and error data. Authentication time did not attain a significant main effect of modality ($F(11,1) = 3.13$, $p = 0.10$) or beat-type ($F(11,1) = 0.29$, $p = 0.6$). The interaction was not significant. Similarly, authentication errors did not yield significant effects of modality ($F(11,1) = 0$, $p = 1$), beat-type ($F(11,1) = 2.07$, $p = 0.17$) or the interaction of these two factors. Resets followed the same pattern: no significant effect on modality ($F(11,1) = 0.58$, $p = 0.46$), beat-type ($F(11,1) = 0.05$, $p = 0.82$) or interaction. Finally, the two-way ANOVA on the TLX overall workload (Fig. 7) also showed no significant effect on modality ($F(11,1) = 0$, $p = 1$) or beat-type ($F(11,1) = 3.34$, $p = 0.09$). However, a more detailed analysis of the TLX showed a significant variation (Table 2) on the perception of performance for beat-type ($F(11,1) = 6.12$, $p = 0.03$), but not for modality ($F(11,1) = 0.49$, $p = 0.49$).

5.4. Discussion

The experimental results show few significant differences. This is an encouraging result, particularly for the modality comparison.

Table 2
TLX metrics and significance values (ANOVA) for Colorlock over beat-type and modality conditions.

	Beat type		Modality		Interaction	
	F	*p*	*F*	*p*	*F*	*p*
ANOVA TLX Colorlock						
Overall workload	3.34	0.094	0	1	0.04	0.84
Mental	3.25	0.098	0.38	0.55	0	1
Physical	0.02	0.89	0.06	0.81	0.01	0.92
Temporal	0.08	3.53	0	1	0.03	0.86
Performance	**6.12**	**0.03**	0.49	0.49	0.03	0.86
Effort	3.49	0.08	0.1	0.75	0.75	0.4
Frustration	1.01	0.33	1.46	0.25	0.86	0.37

The fact that there were no significant changes between the haptic and audio modalities across the full set of measures used in the study (including authentication time, errors, resets and overall workload) strongly indicates equivalent levels of performance across modality, despite literature (Lechelt, 1975) suggesting that audio performs better than haptics in numerosity tasks. This is also in stark contrast to previous PIN entry systems, such as the SpinLock prototype described in this paper, that have included audio and haptic interfaces. Such systems have typically shown haptic cues to be between 16% and 22% slower and more error prone than audio cues (Bianchi et al., 2011b). This suggestion was supported in the post-experiment interviews, in which 33% of participants expressed a preference for the haptic modality, 58% for the audio modality and 8% rated both equally preferred. Taken together, these results strongly suggest that users felt comfortable using both modalities and that the approach to single cue haptic display taken in the Colorlock prototype is highly appropriate to the scenario of PIN entry.

No difference in performance was observed between the two beating types (constant vs. random) deployed in the study. However, the subjective measures revealed higher levels of workload, and specifically lower levels of perceived performance, when exposed to the random beating style. These results were supported by comments reported in the interviews: when asked, all participants indicated that constant beating was easier and consequently preferred.

A more detailed analysis of the 44 incorrectly entered PINs for all the four conditions was also conducted. The mean number of failed authentications per user per condition was 0.91 (SD 1.08,

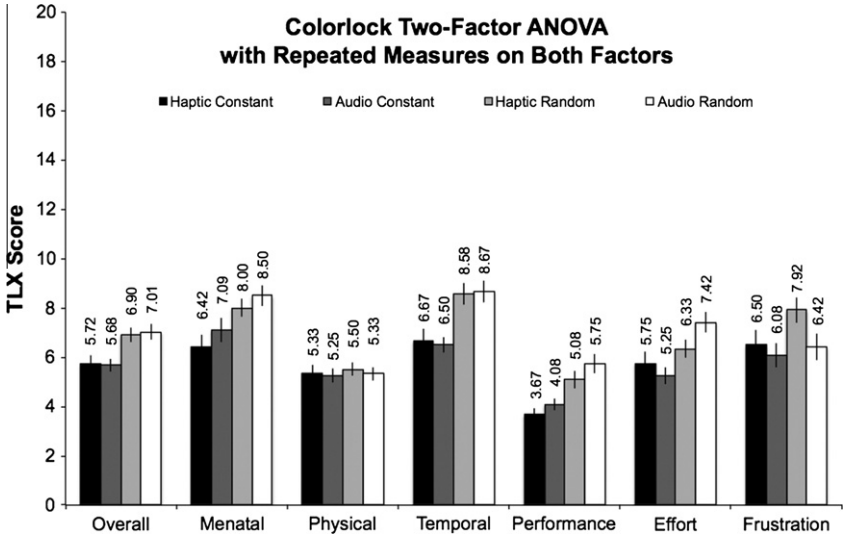

Fig. 7. Colorlock two-factor ANOVA with repeated measures.

median 1, mode 0, maximum 5 minimum 0). The mean number of mistakes per incorrect PIN entry was 1.35 (SD 0.46, median 1.25, mode 1, maximum 4, minimum 1). The majority of errors were due to the selection of the incorrect number of cues (75%) rather than of the incorrect color (25%), and a Pearson test revealed no correlation between the number of errors and the target PIN item ($r = -0.1$, $n = 18$, $p = 0.67$). Incorrectly selected numbers were typically within two digits of the correct item (mean: 1.31, SD: 0.18), with 66% of these errors being one number away from the target. Furthermore, the majority of mistakes (55%, 34 errors) occurred during input of the final PIN item. A likely explanation for this pattern relates to how the PINs are inserted: on entry of the final PIN item, the PIN is marked as complete and the user is unable to use the reset operation to correct any errors that may have occurred. This should be corrected, most simply via the introduction of an explicit enter PIN button, in future iterations on the system design.

5.5. Security evaluation

The security evaluation took two forms. Firstly, a software simulation was developed in order to test the strength of the randomization function. 500 PIN item insertions (100 per possible value) were generated using the constant and random beat algorithms. Pearson's r correlations between time and PIN item were calculated to be 0.84 (constant beat) and 0.87 (random beat), indicating strong relationships. Running the same test on user input during the study led to weaker correlations: 0.75 (constant beat) and 0.77 (random beat). This is most likely due to small temporal irregularities in how and when participants responded to the cues.

Secondly, an observation attack was conducted on video footage of the full 40 authentications that make up the study for three experimental participants. Each participant used the same PIN for the duration of the study, leading to a total of 120 observations of authentication processes, addressing three different PINs and split evenly across all four experimental conditions. The video was shot using a digital camera (60 FPS interlaced) positioned on a tripod behind the user and pointing directly at the device in order to guarantee a clear and unobstructed view of the screen at all times. Participants were instructed that were being filmed and that they should not attempt to obfuscate their fingers or input. The video for each participant was approximately 8 min long. The selection of a subset (25%) of the study as the focus of the observation attack is a common approach in security evaluations (e.g. De Luca et al., 2010; Bianchi et al., 2011c). For the security study described here, a subset of participants, rather than a subset of trials from all participants, was selected to assess the weakness of the system to repetition attack: to a prolonged series of observations of the same user entering the same PIN.

Three security experts performed the observation attack, each taking approximately 2 h to complete the task. To facilitate attackers in their task they were provided with a summary table listing the average time that users needed for inserting each of the five items; attackers did not need to guess or estimate the mean time needed to insert a particular item. Two attackers successfully deduced two of the three PINs, while the third did not correctly ascertain any of the PINs. However, all attackers successfully determined at least two of the PIN items for each of the three PINs. The processes they deployed varied. Two attackers analyzed the video frame-by-frame and manually noted the time-stamp of the user actions so to compare it with the summary table that we provided; the third attacker developed a small piece of software to aid recording multiple time-stamps upon user actions that could be compared with the time stamps in the summary table. Attackers stated this was a long, tedious but not a particularly challenging process. In sum, considering this information in conjunction with the strong correlation between dwell duration and PIN item selected, we conclude that while observation of a PIN entered with Colorlock is more challenging than with a regular keypad, a determined attacker can recover a PIN given a sufficiently large number of repeated observations.

6. Prototype II: Timelock

The Timelock prototype further explores the counting-based single-cue PIN entry interaction model. The interaction strongly resembles that deployed in Colorlock: haptic or audio cues are triggered at constant or random intervals by holding down a virtual button on a touch screen. Timelock extends the Colorlock prototype by introducing a richer range of error-correction mechanisms that allow users to change and correct their data input. This is achieved by combining the input of the PIN items and the display of their state into a single set of manipulable on-screen widgets. Timelock also employs an optimized pair of beat generating random functions intended to maximize temporal performance without increasing error rate. These techniques are described in detail below.

6.1. Design and implementation

In Timelock users input a PIN item by directly touching on-screen widgets that represent the PIN items as four equally sized buttons arranged in a single row (Fig. 8). As with Colorlock, temporally separated non-visual cues are delivered according to the dwell time on these widgets. This direct metaphor for input supports the use of an additional, observable input dimension: the order in which PIN items are entered. For example, although a standard PIN is entered from leftmost item to rightmost item, the Timelock PIN can be entered in any conceivable sequence simply by selecting the PIN item widgets in the desired order. The Timelock system was designed with this functionality in mind: PIN items in the system are restricted to numerical values between one and five, but must also be entered in a predetermined order, substantially increasing the entropy of the system. In fact, the entropy can be calculated as the PIN input range (5) to the power of the number of PIN items (4) multiplied by the factorial of the number of PIN items (4), leading to a final figure of 15,000. This makes the resistance of the Timelock prototype to brute force attack significantly greater than that of a standard purely numerical ATM PIN. Resistance against observation attack is lower, as with Colorlock, and has an entropy of 625 (5^4).

Users are provided with four error correction and recovery mechanisms in Timelock. Firstly, as with the other prototypes described in this paper, users can entirely erase the input PIN via a clear button that triggers a *reset* event. The three new features were inspired by the performance and comments from users during prior evaluations and allow correction of items during an ongoing PIN entry process and without requiring a complete reset. They are: a *back button* that allows users to undo their most recent PIN entry; a *re-entry* mechanism which allows users to re-input the last item they entered by simply selecting and dwelling on the appropriate PIN item button once again and; a *slide-down* technique intended to support correction of errors in which an entered PIN item is one digit greater than that desired. To implement and represent this technique, each PIN item widget functioned not only as a button, but also a "slider" able to move a short distance towards the base of the mobile device screen in response to a downwards swipe over its surface. Once such an operation had taken place, it is visualized by the change in the button position and has the effect of lowering the entered PIN item count by one. For example, dwelling on a PIN item target for a count of four non-visual cues, then performing a downwards swipe over that item would result in a figure of three being entered as the final value of the PIN item.

Please cite this article in press as: Bianchi, A., et al. Counting clicks and beeps: Exploring numerosity based haptic and audio PIN entry. Interact. Comput. (2012), http://dx.doi.org/10.1016/j.intcom.2012.06.005

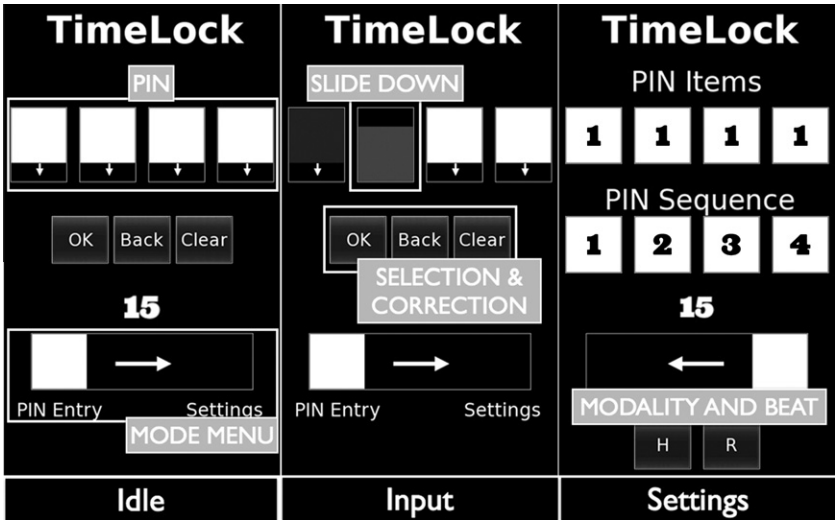

Fig. 8. The Timelock Graphical User Interface: whilst idle (left), during the user interaction (center) and the settings screen showing user password (right).

Cue randomization in Timelock was achieved via improved versions of the *constant beats* and *random beats*. Adjustments were made to the parameters that regulate the initial pause and beating intervals. The changes were as follows: the maximum value of the initial pause was increased to 1500 ms (constant beat) or 2000 ms (random beat) to increase the variability in this first step; the allowable range of inter-cue intervals was decreased to between 300 ms and 400 ms to improve overall temporal performance (Philippi et al., 2008).

The Graphical User Interface used in Timelock is minimal and is again composed of an application screen and a setting screen. The application screen contains the PIN item slots. The slots indicate the current status of the selection (white for no selection, grey for entered selection) and the result of the authentication (green for successfully authentication, red for denied). Moreover, these targets are also the widgets that users hold in order to enter PIN items (via counting non-visual cues), to *re-enter* them, or *slide-down* in order to modify the entered value. Under the PIN slots are buttons for executing the *back* and *reset* operations. There is also a button for completing or entering a finalized PIN. The setting screen contains the widgets necessary to set the PIN values and order, the modality (haptic or audio) and the beat type (constant or random). This prototype was developed, as for Colorlock, for the Samsung Galaxy Tab running the Android operating system.

6.2. Evaluation

A user study modeled on that conducted on the Colorlock prototype was used to evaluate Timelock. Once again, the goals were to compare performance between the two display modalities (haptic vs. audio) and among different beat schemes (constant vs. random beat). Additional points of interest in this prototype were the direct entry of PIN items, the use of PIN item order as a security parameter and the usage rates of the different correction mechanisms (reset, back, slide-down and re-enter).

6.2.1. Participants

The study involved 12 participants (seven male, five female with age between 25 and 33 years, mean: 27.4 and SD: 2.8). They were a mix of students and professionals. Twenty-five percent reported themselves to be advanced computer users, 58% medium and 17% basic. None of the participants were involved in previous related experiments. Participants were compensated for their time with a small gift.

6.2.2. Materials and methods

The materials and methods for the experiment are exactly identical to those used for the Colorlock experiment in terms of condition structure and balancing, practice, total PIN items recorded, duration and use of randomly generated PINs (summary in Table 1). However, in this user study additional measures were taken. Beyond workload, successful PIN entry time, error rate and resets the study also captured the use of the three new correction mechanisms: back, re-entry and slide-down.

6.3. Results

Experimental data are shown in Figs. 9–11. All data were tested using two-way repeated measures ANOVAs. The authentication time did not attain a significant main effect of modality ($F(11,1) = 0.5$, $p = 0.49$), but did attain an effect on beat-type ($F(11,1) = 146.95$, $p < 0.01$). Error rate was not significantly different across modalities ($F(11,1) = 0.45$, $p = 0.51$) nor beat-type ($F(11,1) = 0.14$, $p = 0.71$). ANOVA results for the correction mechanism are reported in Table 3. The slide-down mechanism attained significance for beat-type ($F(11,1) = 5.17$, $p = 0.04$) but not across modality ($F(11,1) = 1.7$, $p = 0.21$), though a significant interaction was found between the two variables ($F(11,1) = 12.16$, $p = 0.01$). An ANOVA over the overall workload (Fig. 12, Table 4) revealed sta-

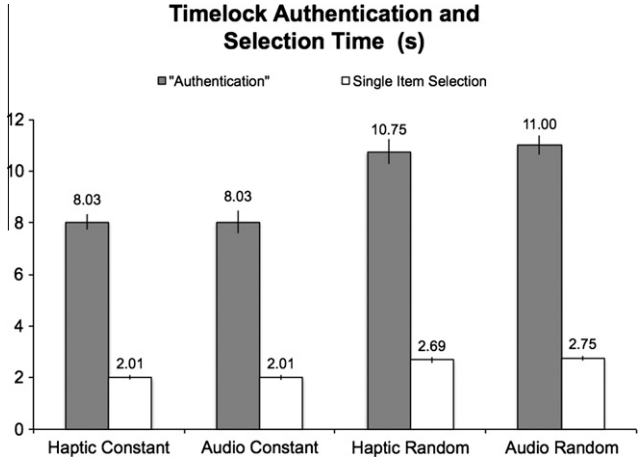

Fig. 9. Timelock authentication and PIN item selection time (seconds).

Please cite this article in press as: Bianchi, A., et al. Counting clicks and beeps: Exploring numerosity based haptic and audio PIN entry. Interact. Comput. (2012), http://dx.doi.org/10.1016/j.intcom.2012.06.005

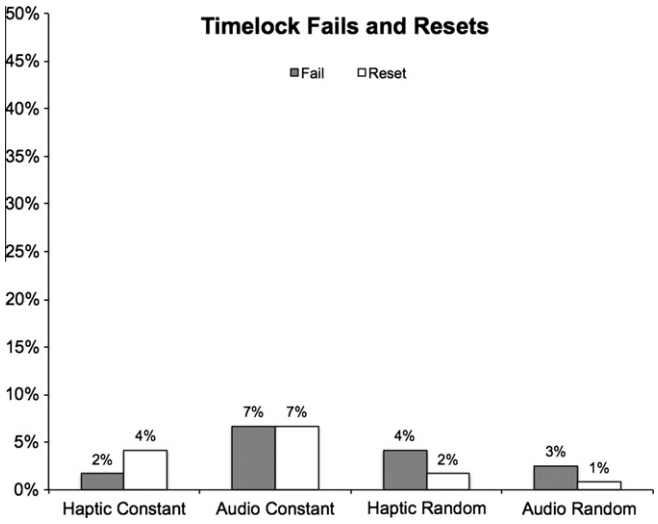

Fig. 10. Timelock fails and resets.

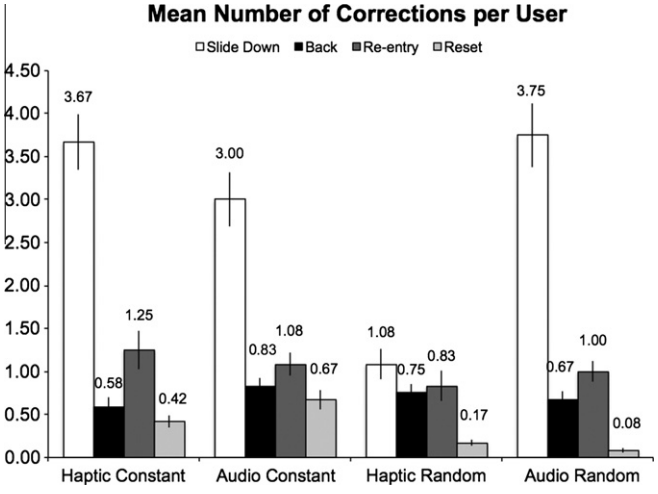

Fig. 11. Timelock mean number of corrections per user (slide down, back, re-entry, reset).

tistical significance between beat-types ($F(11,1) = 5.23$, $p = 0.04$) but not modalities ($F(11,1) = 4.13$, $p = 0.06$). A more detailed analysis of the individual TLX items also showed a significant difference between beat-types for the Physical Demand ($F(11,1) = 7.23$, $p = 0.02$) and Effort Expended ($F(11,1) = 8.75$, $p = 0.01$) and between modalities for Performance Level Achieved ($F(11,1) = 12.88$, $p < 0.01$) and Frustration Experienced ($F(11,1) = 5.73$, $p = 0.03$).

6.4. Discussion

The Timelock results are encouraging. There were no significant differences in performance between audio and haptic modalities, indicating that participants were equally able to use the system with cues in either modality. This result was reinforced by the TLX data, which showed no difference in overall subjective workload between audio and haptic conditions. Therefore, the Timelock prototype is able to present haptic cues, which are typically considered to be significantly less rich that audio cues, in a manner that yields equivalent performance and perceived workload.

Significant differences were observed between the two beat-types used in the study, both in terms of the time to complete the trials and in the frequency with which the slide-down error correction mechanism was used. The first result is directly due to the parameters defining the ranges of the initial pause and inter-cue intervals. The second result may be due to increased levels of attention – the TLX data suggests that participants found the random condition more challenging than the constant condition.

Table 3
TLX metrics and significance values (ANOVA) for Timelock over beat-type and modality conditions.

	Beat type		Modality		Interaction	
	F	*p*	*F*	*p*	*F*	*p*
ANOVA TLX Timelock						
Overall workload	**5.23**	**0.04**	4.13	0.06	0.32	0.58
Mental	2.48	0.14	1.13	0.31	0.46	0.51
Physical	**7.23**	**0.02**	0.18	0.67	0.31	0.58
Temporal	2.75	0.12	2.4	0.14	0.01	0.92
Performance	0.22	0.64	**12.88**	**0.01**	0.37	0.55
Effort	**8.75**	**0.01**	2.76	0.12	0.26	0.62
Frustration	3.58	0.08	5.73	**0.03**	0	1

The temporal performance with Timelock compares favorably to that attained in studies of the other prototypes discussed in this paper. Overall mean authentication time across the study was 9.45 s (SD 1.6 s), while in the optimal condition, featuring the constant beat-type and both haptic and audio feedback, mean authentication times were 8 s. This represents an improvement in performance of approximately 5% from that attained in Colorlock and 38% from that attained in Spinlock. Error and reset rates were also reduced compared to the prior prototypes. Respectively they counted only for 4% and 3.5% of the total number of authentications, figures that are very low for experimental settings and which were stable and unchanging across all four experimental conditions. Indeed across the whole study, the mean number of failed authentications was very low: 0.3 (SD 0.9, median 0, mode 0, maximum 5, minimum 0). These figures also improved on the those reported for the both Colorlock prototype (improvements of 45% in error rates and 76% in reset rates) and the Spinlock prototypes (23% lower error rate and 95% reduced resets).

An in-depth analysis of incorrect trials indicates 90% were due to errors in counting the cues displayed for each PIN item and only 10% due to selecting PIN items in an incorrect order. Furthermore, the mean number of incorrect PIN items per failed authentication was 1.03 and 73% of those errors involved the selection of a number of cues only one away from the correct target (51% involved overshooting the target by one and 22% undershooting it by one). Although this data documents error trials, it demonstrates that exceeding the desired count target by one item is a common problem, and therefore justifies the inclusion of the slide-down correction mechanism.

Participants' usage patterns of the error correction mechanisms during successful trials, shown in Fig. 13, reinforces this point. The slide-down technique was most commonly used (56%), followed by the PIN re-entry (22%), the back button (16%) and finally resets (7%). These data correspond with comments captured from the participants in the post hoc interview: 58% of them stated a preference for the slide-down technique while only the 8% preferred the reset button. The correction mechanisms were also employed highly accurately. In 84% of uses, the correction mechanisms were deployed to adjust an incorrectly entered data item to the appropriate item; in 11% they were used on an incorrect item, but the adjustment was not successful. Reasons for an incorrect adjustment were predominantly failures to successfully execute the downwards stroke required to make a slide-down correction. In only 5% of cases were the correction mechanisms erroneously deployed to adjust a correct PIN item. Together these results suggest two things. Firstly, that the correction mechanisms introduced in

Please cite this article in press as: Bianchi, A., et al. Counting clicks and beeps: Exploring numerosity based haptic and audio PIN entry. Interact. Comput. (2012), http://dx.doi.org/10.1016/j.intcom.2012.06.005

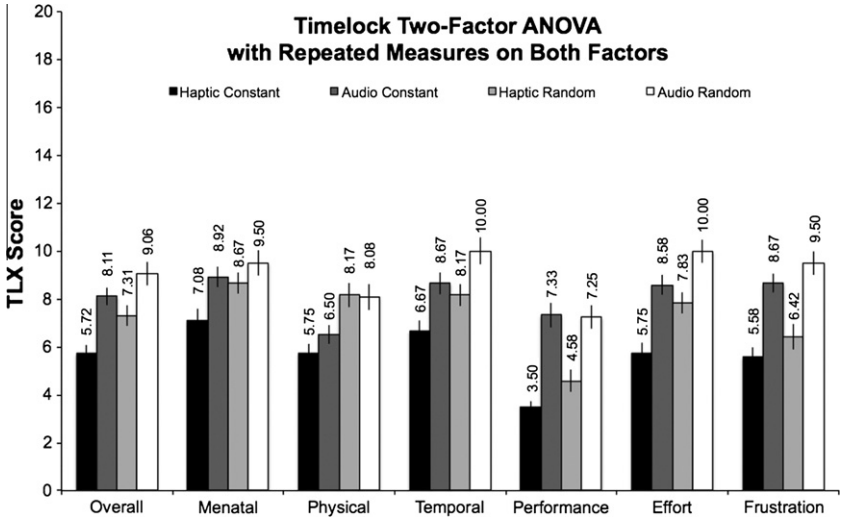

Fig. 12. Timelock two-factor ANOVA with repeated measures.

Table 4
Timelock ANOVA for the correction mechanisms.

	Beat type		Modality		Interaction	
	F	*p*	*F*	*p*	*F*	*p*
ANOVA Timelock correction						
Slide down	**5.17**	**0.04**	1.7	0.21	**12.16**	**0.01**
Back	0	1	0.22	0.64	0.18	0.67
Re-entry	0.2	0.66	0	1	0.17	0.68
Resets	2.57	0.13	0.15	0.7	1	0.33
Average across conditions	3.69	0.08	1.57	0.23	2.73	0.12

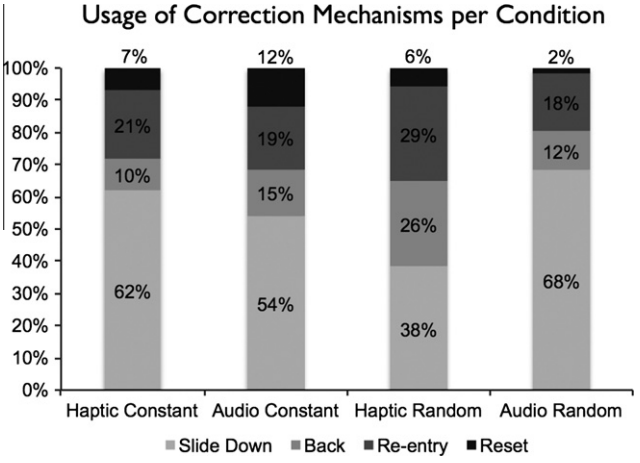

Fig. 13. Timelock usage of correction mechanisms per conditions.

Timelock are effective at reducing the overall number of errors and the use of resets operations, making PIN entry a more efficient, robust and ultimately less frustrating experience. Secondly, the fact that users were able to rapidly, reliably and accurately deploy these error correction techniques strongly indicates that they were confident in their numerosity judgements. This represents a strong endorsement of the counting approach deployed throughout this paper and is particularly powerful as it contrasts with tactons and earcons recognition rates of about 70% (Brewster et al., 1993; Brown et al., 2005), levels at which participant confidence in the correctness of their responses is likely to be relatively low.

Further support for the system came in the form of subjective comments made during the post-experiment interview. In this session, 83% of participants declared that the PIN was easy to remember, with 75% preferring the constant beat, 8% the random beat and 17% expressing no preference. Finally, 75% of users expressed a preference for the haptic condition, a figure supported by detailed analysis of the TLX questionnaire, in which the individual factors of Performance level achieved and Frustration experience were improved in the Haptic condition. This strong endorsement of haptics is in contrast to preference data collected in the other studies reported in this paper.

6.5. Security evaluation

Timelock implements optimized versions of the constant and random beat functions; a process similar to that deployed in Colorlock was used to evaluate the security of the interface. 500 samples (100 for each PIN item) were again generated and Pearson's *r* between time and PIN item was 0.7 for the constant beat and 0.62 for the random beat, indicating weaker relationships than in the Colorlock prototype. This observation was verified in a simulated observation attack on the full set of 40 PIN entries performed by three of the experimental participants. The attacks were conducted by the same three security experts as in the Colorlock evaluation; once again, each attacker spent approximatively 2 h in total and deployed similar methods. For Timelock, none of the attackers successful deduced a PIN, and on average correctly guessed only 1.2 (SD 0.8) items per PIN. The attack was reported to be much more challenging than for Colorlock, due to the higher variance of the collected time stamps and the usage of correction mechanisms which made observation more difficult. In conclusion, this preliminary security suggests that changes to the interface and optimization of the beating functions resulted in a substantially more secure interface in which attackers were unable to deduce PINs even when provided with unlimited time and a video of 40 repeated PIN entry processes. This preliminary finding suggests that the technique is sufficiently resistant to observation to be of practical use.

7. Overall discussion

This paper summarized the design and study of SpinLock, a PIN entry system using a cue counting input mechanism and based on the metaphor of the dial lock of a safe, and introduced, described and evaluated Colorlock and Timelock, two techniques which fur-

ther explore the use of cue counting techniques based on target selection and dwell-times. Contrasting among the performance reported in these three systems provides some valuable lessons. Perhaps the clearest distinction lies in the clear superiority of Colorlock and Timelock over the Spinlock prototype. The former two interfaces average mean PIN entry times of 9.7 s, error rates of around 5.6% and ratings of TLX overall workload of 6.9; the latter resulted in times of 15.4 s, error rates of 6% and workload reported at 9.1. These improvements are most likely due to the shift from an input technique based on a continuous gesture (the rotational movement in SpinLock) to one based on simple target dwell times (in Colorlock and Timelock) and strongly advocate for the use of such elementary input actions during display of non-visual, and particularly haptic, cues. Some evidence to support this assertion comes from previous work suggesting that haptic cues are harder to accurately sense whilst users are engaged in physical tasks (Oakley and Park, 2008) – perhaps unsurprisingly, the act of physical motion interferes with haptic perception. Consequently, interfaces that minimize this overlap and disturbance can lead to higher levels of performance.

A second key observation from the set of studies is that, in the Colorlock and Timelock systems, there were few differences in performance between audio and haptic modalities. This is unusual. Indeed, in research considering equivalent haptic and audio tasks (e.g. Bianchi et al., 2011a) performance typically resembles that observed in the SpinLock prototype: audio task times, error rates and workload are considerably reduced compared to those observed in haptic conditions. This finding is supported by the general literature on numerosity, which suggests it can be performed equivalently across audio and haptic modalities (White and Cheatham, 1959). More practically, it also suggests that cue-counting may be a good way to create other tactile and haptic interfaces that are clear, unambiguous and reliable.

Data describing how the Timelock corrections mechanisms were used further support this claim. In only 5% of cases were they inappropriately deployed to adjust correct data input, suggesting that the vast majority of errors take the form of physical slips and the errors emerging from the varying inter-cue intervals rather than more serious cognitive mis-counts.

The three interfaces also explored the combination of different forms of standard input, in the form of visually observable gestures or buttons presses, and non-visual information presentation. The lack of errors due to incorrect visual input (or subjectively reported confusion) throughout the three evaluations suggest that the combinations deployed were simple and effective – users were able to understand the combinations of gestures or button presses and non-visual cues with little difficulty. However, in terms of preventing an observation attack, the proportion of observable to non-observable information is an important one. Each of the systems exceeded an entropy of 10,000 (standard for a 4 digit ATM PIN) against brute force attacks, but achieved considerably greater resistance to observation attack – an entropy of 625 – than that offered by standard keypads.

It is worth unpacking this comparison in detail. A wide range of literature has been reported investigating performance using standard textual passwords and numerical PINs, the oldest and most common authentication methods on both public terminals and mobile devices. For instance, De Luca et al. (2010) recently reported zero errors for four digit PIN entry at a public terminal. Input times were between 1.32 s (SD 0.8 s) and 1.56 s (SD 0.37 s), depending on whether the PIN was set by a user or randomly generated. However, resistance to observation was low: simple in-person shoulder-surfing led to attackers correctly acquiring PINs in 77% of attempts. Similarly, Lee and Park (2005) evaluated PIN entry on mobile phones in a lab study. They showed that, in the absence of training, input time is 2.4 s (SD 0.4) and error rate 9.4% (SD 1.2). Improvements after training lead to input times of 1.6 s (SD 0.3) and error rate of 5.2% (SD 1.7). Resistance to observation remained very low: 92.4% (SD 1.8) of shoulder surfing attempts were successful.

This paper argues that the three prototypes it describes considerably improve on this performance and uses security studies to validate this claim. In particular this investigated a potential security weakness of the cue-counting approach: counts inherently increase with time or distance and may therefore be inferred from observation of user input. To explore this issue, the three systems described in this paper introduced different forms of randomness into the cue presentation sequences. Results indicate that the optimal combination of user performance and resistance to observation appear to be achieved using *constant* beat algorithms which deploy an initial pre-cue pause which can span a relatively large randomly determined range followed by beats issued at regular, equally spaced intervals. The security attacks reported in this paper provide a preliminary test of the security level of these prototypes and indicate that, though attackers might eventually be able to deduce a PIN from observing and recording multiple inputs, the process is far more difficult and cumbersome than with standard ATM input techniques (De Luca et al., 2010).

It is also valuable to contextualize the results reported in this paper against previous research on non-visual PIN entry techniques: Table 5 contains a summary of this information. Although diversity in methodologies, conditions and participants make formal comparison impossible, a qualitative interpretation is very supportive: it suggests the approaches and designs presented in

Table 5
Haptic and audio PIN entry technique in comparison.

Name	Security brute force	Security observation	Technique	Modality	Time (s)	Errors
4-digit PIN (keypad) (De Luca et al., 2010; Lee and Park, 2005)	1 in 10,000	No security	Unimodal	Vision	~1.5	~0
Undercover (Sasamoto et al., 2008)	1 in 10,000 or less	1 in 10,000 or less	Multimodal recognition	Haptic + vision	~25–45	~26–52
Vibrapass (De Luca et al., 2009)	1 in 10,000 or less	1 in 10,000 or less; weak against two or more	Multimodal direction	Haptic + vision	~6–19	8
Tactile authentication (Kuber and Yu, 2010)	1 in 6561 of less	1 in 6561 or less	Unimodal recognition	Haptic	~38	~6
Haptic keyboard (Bianchi et al., 2010a)	1 in 10,000 or less	1 in 10,000 or less	Unimodal recognition	Haptic	33.8	6.7
Haptic wheel (Bianchi et al., 2010b)	1 in 10,000 or less	1 in 10,000 or less	Unimodal recognition	Haptic	23.2	16.4
Phone lock (Bianchi et al., 2008)	1 in 10,000 or less	In 10,000 or less	Unimodal recognition	Haptic	19.9	6.6
			Unimodal recognition	Audio	12.2	4.7
Spinlock Bianchi et al., 2011b	1 in 10,000 or less	1 in 10,000 or less	Unimodal counting	Haptic	13.8	8.3
			Unimodal counting	Audio	10.8	3.3
Colorlock	1 in 10,000 or less	1 in 625	Unimodal counting	Haptic	~10	7
			Unimodal counting	Audio	~10	7.5
Timelock	1 in 10,000 or less	1 in 625	Unimodal counting	Haptic	~8	2
			Unimodal counting	Audio	~8	7

Please cite this article in press as: Bianchi, A., et al. Counting clicks and beeps: Exploring numerosity based haptic and audio PIN entry. Interact. Comput. (2012), http://dx.doi.org/10.1016/j.intcom.2012.06.005

this paper were both fast and accurate. We provide two explanations for this. Firstly, we suggest that the uni-modal nature of the interactions studied in this paper is an important contributor to the speed and accuracy with which they were performed. In many scenarios, the use of multi-modal cues divides central cognitive resources such as attention and consequently results in low levels of performance (Spence and Driver, 1997). Secondly, we argue that numerosity is an effective approach to non-visual interaction in high bandwidth, focused data entry tasks; in many scenarios it may be more effective than traditional recognition approaches based on earcons, audio icons or tactons. Direct support for these claims comes in the form of data reported in related literature. For example, evaluations of Sasamoto's multi-modal (haptics plus visuals) PIN entry system Undercover (Sasamoto et al., 2008) led to mean authentication times of ~25–45 s and error rates of ~26–52%. Similarly, PhoneLock, a uni-modal recognition based PIN entry system (based on tactons or audio icons) achieved mean authentication times and error rates of 18.8 s and 7%. Moving beyond PIN entry tasks and considering the general literature on recognition of audio and haptic cue sets, there are many studies focused on reporting error rates for distinguishing individual cues from sets of about five audio or haptic stimuli. Recognition rates for haptics and audio are about 70% with sets of five or less stimuli (tactons or earcons) (Brown et al., 2005). In contrast, the cue counting technique used in the studies described here enables, through rapid, repeated interaction, accurately selecting a specific set of cues from a possible space of thousands of possible PINs in times as low as 8 s.

In summary, the work in this paper demonstrates the design of a cue counting technique in the focused, high demand scenario of PIN-entry. It serves as a practical example of how audio and haptic cue counting techniques can be successfully applied to challenging, focused cognitive tasks. We argue that the techniques it advocates are applicable to other high-entropy non-visual interaction tasks that require attention and demand accuracy during use. These may include existing application areas for non-visual interaction such as automative UIs (Chang et al., 2008) and surgical systems (Boulanger et al., 2006).

8. Limitations and future work

The objective performance in the non-visual systems discussed in this paper is poor when compared with traditional numerical PIN entry systems; at least six times worse (De Luca et al., 2010; Lee and Park, 2005). This reduction in performance is counter-balanced by a substantial increase in the resistance to observation attacks in non-visual PIN entry; its clear that keypad techniques have little or no resistance to observation either in person or via cameras. This suggests that non-visual PIN-entry may be a viable solution only for applications requiring increased levels of security, be it for restricting access to storage systems, physical spaces, or digital resources. Examples of devices currently targeting this area include the electronic combination locks developed by KABA MAS corporation (X-09, Dawson et al., 2004). Given that such systems already require complex, lengthy authentication processes, non-visual interaction may be able to provide improved security with minimal costs to objective performance. Explicit investigations of the application of non-visual cues to such devices is an obvious next step for this work.

The techniques may also be applicable to other user tasks. Previous research indicates that users deal with large numbers of PINs and passwords for a wide range of tasks on a daily basis and that this content is presented using a diverse set of modalities and input techniques (Grawemeyer and Johnson, 2011). One concrete example is the task of securely pairing two devices (Ion et al., 2010) which can achieved via PIN entry or more experimental techniques such as taking a picture of a barcode (Ion et al., 2010). This research suggests that users select different methods according to demands of particular situations. Data from a recent diary study also reports that authentication is increasingly happening in a wide range of situations: 40.8% of authentications take place outside of the home and 33.6% involve a device that is not a personal computer or mobile phone (Hayashi et al., 2011). These findings indicate that there may be niches where users are willing to trade the increased temporal burdens of non-visual authentication for its improved level of security.

This work is also limited by the fact that it is lab-based; the studies include a relatively small number of demographically homogeneous participants and take place over a short period of time in controlled settings. These factors limit the confidence with which we can generalize the findings and clearly extending the work to include longitudinal studies involving a more diverse participant pool in more realistic scenarios will effectively complement the current paper. This work is also limited by its focus on a specific application task. Although inspired by the literature on human processing of non-visual cues, it moves substantially beyond these foundations. Consequently, fundamental issues of memorization, recall and communication of passwords involving non-visual cues are not fully understood nor considered in this paper. Work to better understand these cognitive process would support the design of cue sets and techniques that can be used more rapidly and reliably. In particular, optimizing haptic and audio cue design and spacing interval to both maximize the resistance to observation attack and also the ease with which counting (or ideally subitising, Gallace et al., 2008) operations can take place is a key area for future study. Another important topic relates to PIN memorization and how users are able to deploy techniques such as mnemonics, associations or external aids to help them achieve this. Memory is a critical aspect of any password system, including those using non-visual cues (e.g. Kuber and Yu, 2010; Yan et al., 2004) and future extensions to this work need explicitly consider it.

There are also considerable limitations to the security evaluations presented in this paper. The attacks conducted were highly specific to a threat model composed of either brute-force attack or repeated visual observation, either in person or via camera. While the performance against these vectors was strong, a wide range of other methods also demand attention in future research. These include observation in non-visual modalities such as by directional microphones, vibration sensors, or techniques that detect the variations in electromagnetic fields caused by motor activation. As with much security research, it would also be valuable to consider attacks based on physical interventions, such as tampering with or stealing a device, cross-cutting social attacks such as phishing, and the susceptibility of the approach to malicious software such as keyloggers. Equally, future work should develop and evaluate counter-measures to these attacks. For instance, in the context of non-visual observation attack, existing techniques for masking auditory signals (Gescheider, 1966, 1967) or the sounds generated by haptic apparatus will likely be useful avenues of investigation.

9. Conclusions

In conclusion, this paper introduced a novel unimodal non-visual interaction technique for PIN entry at public terminals, such as bank ATMs or door-locks, that is both secure (e.g. resistant to brute force and observation attacks) and usable. The proposed technique is based on temporal numerosity, the human ability to accurately, confidently and rapidly count the number of cues presented in rapid temporal succession. This paper summarizes one

Please cite this article in press as: Bianchi, A., et al. Counting clicks and beeps: Exploring numerosity based haptic and audio PIN entry. Interact. Comput. (2012), http://dx.doi.org/10.1016/j.intcom.2012.06.005

system instantiating this concept and introduces two others. Results from the user studies of these three systems show considerable improvements over previous work based on both multimodal or recognition systems, and suggest that this technique is feasible for a range of applications in the security domain. Moreover, the success achieved in the system implies that cue counting approaches to non-visual interaction may be useful in a wide range of applications currently reliant on cue recognition; cue-counting may be faster and less error prone than distinguishing individual cues from a predetermined set. In sum, this paper introduces a novel non-visual input technique based on numerosity and provides empirical evidence indicating users attain high levels of performance with it in PIN entry, a precise and demanding task. This is a promising initial result and future work should explore and explain the application of this technique to a range of other tasks.

Acknowledgements

We thank Andrew Zoz Brooks for his support during the security studies, all our experimental participants for their time and our reviewers for their valuable comments.

References

Banala, S.K., Seok, H.K., Agrawal, S.K., Scholz, J.P., 2008. Robot assisted gait training with active leg exoskeleton (ALEX). Biomedical Robotics and Biomechatronics, BioRob 2008, pp.653–658.

Bau, O., Poupyrev, I., Israr, A., Harrison, C., 2010. TeslaTouch: electrovibration for touch surfaces. In: Proceedings of UIST '10. ACM, New York, NY, pp. 283–292.

Bernhard, N., Stephan, D., Lutz, E., Andreas, K., 2009. The new BMW iDrive – applied processes and methods to assure high usability. LNCS 5620, 443–452.

Bianchi, A., Oakley, I., Kwon, D.S., 2010a. The secure haptic keypad: design and evaluation of a tactile password system. In: Proceedings of CHI 2010. ACM, New York, NY, pp. 1089–1092.

Bianchi, A., Oakley, I., Lee, J.K., Kwon, D.S., 2010b. The haptic wheel: design and evaluation of a tactile password system. In: Proceedings of CHI 2010. ACM, New York, NY, pp. 3625–3630.

Bianchi, A., Oakley, I., Kostakos, V., Kwon, D.S., 2011a. The phone lock: audio and haptic shoulder-surfing resistant PIN entry methods. In: Proceedings of ACM TEI 2011. ACM, New York, NY, pp. 197–200.

Bianchi, A., Oakley, I., Kwon, D.S., 2011b. Spinlock: a single-cue haptic and audio PIN input technique for authentication, haptic and audio interaction design (HAID 2011). Lecture Notes in Computer Science (LNCS) 6851, 81–90.

Bianchi, A., Oakley, I., Kwon, D.S., 2011c. Using mobile device screens for authentication. In: Proc. of OzCHI 2011. ACM, New York, NY, USA, pp. 50–53.

Biggs, J., Srinivasan, M.A., 2002. Haptic interfaces. In: Stanney, K. (Ed.), Handbook of Virtual Environments. Lawrence Erlbaum, pp. 93–116.

Boulanger, P., Wu, G., Bischof, W.F., Yang, X.D., 2006. Hapto-audio-visual environments for collaborative training of ophthalmic surgery over optical network. Haptic Audio Visual Environments and their Applications, 2006, HAVE 2006 2006, 21–26.

Brewster, S.A., Brown, L.M., 2004. Non-visual information display using tactons. In: Proceedings of CHI 2004 Extended Abstracts, pp. 787–788.

Brewster, S.A., Wright, P.C., Edwards, A.D.N., 1993. An evaluation of Earcons for use in auditory human–computer interfaces. In: Proceddings of InterCHI 1993. ACM, New York, NY, USA, pp. 222–227.

Brown, L.M., Brewster, S.A., Purchase, H.C., 2005. A first investigation into the effectiveness of tactons. In: Proceedings of the First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC '05), pp. 167–176.

Chang, A., Gouldstone, J., Zigelbaum, J., Ishii, H., 2008. Pragmatic haptics. In: Proceedings of TEI '08. ACM, New York, NY, pp. 251–254.

Dawson, G.L., Thompson, D.L., Stallard, L., 2004. United States Patent 6741160. High Security Electronic Combination Lock.

De Luca, A., von Zezschwitz, E., Hußmann, H., 2009. Vibrapass: secure authentication based on shared lies. In: Proceedings of CHI 2009. ACM, New York, NY, pp. 913–916.

De Luca, A., Hertzschuch, K., Hussmann, H., 2010. ColorPIN: securing PIN entry through indirect input. In: Proceedings of CHI 2010. ACM, New York, NY, USA, pp. 1103–1106.

Gallace, A., Tan, H.Z., Spence, C., 2007. The body surface as a communication system: the state of the art after 50 years. Presence: Teleoperators and Virtual Environments 16 (6), 655–676.

Gallace, A., Tan, H.Z., Spence, C., 2007. Multisensory numerosity judgments for visual and tactile stimuli. Perception and Psychophysics 69 (4), 487–501.

Gallace, A., Tan, H.Z., Spence, C., 2008. Can tactile stimuli be subitised? An unresolved controversy within the literature on numerosity judgments. Perception 37 (5), 782–800.

Gaver, W., 1989. The SonicFinder: an interface that uses auditory icons. Human–Computer Interaction 4 (1), 67–94.

Gescheider, G.A., 1966. The resolving of successive clicks by the ears and skin. Journal of Experimental Psychology 71, 378–381.

Gescheider, G.A., 1967. Auditory and cutaneous temporal resolution of successive brief stimuli. Journal of Experimental Psychology 75, 570–572.

Grawemeyer, B., Johnson, H., 2011. Using and managing multiple passwords: a week to a view. Interacting with Computers 23 (3), 256–267.

Hart, S.G., Staveland, L.E., 1988. Development of a multi-dimensional workload rating scale. In: Human Mental Workload. Elsevier, pp. 139–183.

Hayashi, E., Hong, J., 2011. A diary study of password usage in daily life. In: Proceedings of CHI 2011. ACM, New York, NY, USA, pp. 2627–2630.

Henriques, D.Y.P., Soechting, J.F., 2005. Approaches to the study of haptic sensing. Journal of Neurophysiology 93 (6), 3036–3043.

Ho, C., Tan, H.Z., Spence, C., 2005. Using spatial vibrotactile cues to direct visual attention in driving scenes. Transportation Research Part F: Traffic Psychology and Behaviour 8 (6), 397–412.

Hoggan, E., Brewster, S., 2007. Designing audio and tactile crossmodal icons for mobile devices. In: Proceedings of the 9th International Conference on Multimodal interfaces (ICMI 2007). ACM, New York, NY, USA, pp. 162–169.

Ion, I., Langheinrich, M., Kumaraguru, P., Capkun, S., 2010. Influence of user perception, security needs, and social factors on device pairing method choices. In: Proceedings of the Sixth Symposium on Usable Privacy and Security (SOUPS 2010). ACM, New York, NY, USA (Article 6).

Kortum, P., 2008. HCI Beyond the GUI: Design for Haptic, Speech, Olfactory, and Other Nontraditional Interfaces (Interactive Technologies). Morgan Kaufmann Publishers Inc., San Francisco, CA.

Kuber, R., Yu, W., 2010. Feasibility study of tactile-based authentication. International Journal of Human–Computer Studies 68 (3), 158–181.

Lechelt, E.C., 1975. Temporal numerosity discrimination: intermodal comparison revisited. British Journal of Psychology 66, 101–108.

Lee, S.J., Park, S.B., 2005. Mobile password system for enhancing usability-guaranteed security, HSI 2005. LNCS 3597, 66–74.

MacLean, K., 2003. Perceptual design of haptic icons. In: Proceedings of EuroHaptics 2003, IEEE 2003.

Malek, B., Orozco, M., Saddik, A., 2006. Novel shoulder-surfing resistant haptic-based graphical password. In: Proceedings of EuroHaptics, Paris, France, July 2006.

Oakley, I., Park, J., 2008. Did you feel something? Distracter Tasks and the Recognition of Vibrotactile Cues, Interacting with Computers 20 (3), 354–363.

Oviatt, S., 1999. Ten myths of multimodal interaction. Communication of ACM 42 (11), 74–81.

Philippi, T., van Erp, J.B.F., Werkhoven, P.J., 2008 Multisensory temporal numerosity judgment. Brain Research 1242, 116–125.

Pirhonen, A., Tuuri, K., 2008. In Search for an integrated design basis for audio and haptics, haptic and audio interaction design. LNCS, pp. 81–90.

Qian, H., Kuber, R., Sears, A., 2009. Towards identifying distinguishable tactons for use with mobile devices. In: Proceedings of the ACM SIGACCESS 2009, pp. 257–258.

Reiner, M., 2004. The role of haptics in immersive telecommunication environments. IEEE Transactions on Circuits and Systems for Video Technology 14 (3), 392–401.

Robles-De-La-Torre, G., 2006. The importance of the sense of touch in virtual and real environments. IEEE Multimedia 13 (3), 24–30 (special issue on Haptic User Interfaces for Multimedia Systems).

Sasamoto, H., Christin, N., Hayashi, E., 2008. Undercover: authentication usable in front of prying eyes. In: Proceedings of CHI 2008. ACM, New York, NY, pp. 183–192.

SHAKE SK6. <http://code.google.com/p/shake-drivers>.

Spence, C., Driver, J., 1997. Cross-modal links in attention between audition, vision, and touch: implications for interface design. International Journal of Cognitive Ergonomics 1 (4), 351–373.

Spence, C., Nicholls, M.E.R., Driver, J., 2001. The cost of expecting events in the wrong sensory modality. Perception and Psychophysics 63, 330–336.

ten Hoopen, G., Vos, J., 1979. Effect on numerosity judgment of grouping of tones by auditory channels. Attention, Perception, and Psychophysics. 26 (5), 374–380.

Ternes, D., Maclean, K.E., 2008. Designing Large Sets of Haptic Icons with Rhythm, EuroHaptics 2008. Springer-Verlag, pp. 199–208.

van Erp, J., 2002. Guidelines for the use of vibro-tactile displays in human computer interaction. In: Proceedings of EuroHaptics 2002.

White, C.T., Cheatham, P.G., 1959. Temporal numerosity: IV. A comparison of the major senses. Journal of Experimental Psychology 58 (6), 441–444.

Wickens, C.D., 2008. Multiple resources and mental workload. Human Factors: The Journal of the Human Factors and Ergonomics Society 50 (3), 449–455.

Yan, J., Blackwell, A., Anderson, R., Grant, A., 2004. Password memorability and security: empirical results. Security and Privacy, IEEE 2 (5), 25–31.