VIP: a visual approach to user authentication

Antonella De Angeli, Mike Coutts, Lynne Coventry & Graham I. Johnson
NCR - FSD, Self Service Strategic Solutions
Discovery Centre, 3 Fulton Road DD2 4SW, Dundee -UK
David Cameron & Martin H. Fischer
University of Dundee, Dept. of Psychology
{Antonella.De_Angeli, Mike.Coutts, Lynne.Coventry, Graham.Johnson} @ncr.com

ABSTRACT
This paper addresses knowledge-based authentication systems in self-service technology, presenting the design and evaluation of the Visual Identification Protocol (VIP). The basic idea behind it is to use pictures instead of numbers as a means for user authentication. Three different authentication systems based on images and visual memory were designed and compared with the traditional Personal Identification Number (PIN) approach in a longitudinal study involving 61 users. The experiment addressed performance criteria and subjective evaluation. The study and associated design exploration revealed important knowledge about users, their attitudes towards and behaviour with novel authentication approaches using images. VIP was found to provide a promising and easy-to-use alternative to the PIN. The visual code is easier to remember, preferred by users and potentially more secure than the numeric code. Results also provided guidelines to help designers make the best use of the natural power of visual memory in security solutions.

Categories and Subject Descriptors
H.5.2. [**User Interfaces**]: *Ergonomics, prototyping, and user-centered design.*

General Terms
Design, Security, Human Factors,

Keywords
User authentication, visual memory, security, usability.

1. INTRODUCTION
User authentication is a central component of secure systems that provide access to confidential information or offer personalised services. Historically, methods of establishing the identity of an unknown person have relied either upon an object which they uniquely should possess (token-based authentication) and/or on some secret knowledge, which they uniquely should have (knowledge-based authentication). Special tokens date back to the Bronze Age and passwords to the Roman Centurions [3].

User authentication in computer systems is normally achieved by knowledge-based techniques. People identify themselves by providing a unique user identifier (ID), which they then authenticate with a password. Current Automatic Teller Machines (ATMs) require a combination of a token (bankcard) and secret knowledge (Personal Identification Number or PIN). Other approaches, such as biometrics verification, are being investigated with a view to their general introduction, either in combination with or as a replacement for card and PIN solutions. Biometrics systems make use of anatomical, physiological or behavioural characteristics of an individual for identification or verification purposes [3]. These systems may be a future solution for self-service technology, but there are still many issues with respect to adopting them in the ATM environment. The principal ones are the difficult trade-off between false accept and false reject rates, as well as the storage and handling over the network of biometrics templates. In addition, biometrics systems require specific devices, some of which may be difficult to use or inappropriate for the ATM environment.

Knowledge based-authentication systems are still the pervasive solution. Many people use PINs and passwords for a multitude of devices, from the car radio and mobile phone, to the computer and their bank information. Nevertheless, passwords and PINs have a number of well-known deficiencies reflecting a difficult compromise between security and memorability [1, 2, 5, 14]. A maximally secure password corresponds to a random selection of an alphanumeric string being as long as the system allows. The human limitation on precise recall of meaningless materials is in direct conflict with this requirement. Strict password policies, such as forcing users to change them periodically and to use different codes for different services, make passwords even more difficult to remember. It has been noted that if people are permitted to choose their own passwords they tend to choose ones that are related to their everyday life and which can easily be guessed [1, 7, 14]. Also, people are often lax about the security of this information and may deliberately share the information, or record the PIN and even keep it with the card itself.

So far, the majority of solutions to the problem of weak passwords and PINs have been very technical in nature. A number of proactive measures have been developed to identify weak passwords before they are broken or to increase the computational overhead of cracking programs. Encryption methods and transmission protocols are improving continuously but current authentication systems still suffer from a general neglect of human factors. So far, the proactive actions addressing the user have exclusively regarded training and education. These actions are aimed at raising security awareness and developing strategies to motivate the users to behave in a secure manner [2]. An interesting proposal is the *pass phrase approach* to password generation [14]. It suggests creating a simple sentence of 8 words and choosing letters form these words to generate a secure and memorable password.

NCR is investigating different approaches to cope with the human constraints in the security chain. The aim is facilitating user authentication in public technology, which mediates access to personal bank accounts. This is a highly constrained environment with strong usability and security issues. Consumers of all types need to 'walk up and use' the same machine engaging in a very brief goal-oriented and secure interaction. Owners of terminals cannot allow the ATM to be an easy target for fraud but they cannot afford customer dissatisfaction through false rejection either. The goal of our research is finding the best compromise between usability, cognitive constraints of human memory, and security in authentication systems for accessing ATMs.

This paper reports on our experience with the Visual Identification Protocol (VIP), a project aimed at improving user authentication in self-service technology by replacing the precise recall of a numerical code with the recognition of previously seen images, a skill at which humans are remarkably proficient. It describes how the idea of using pictures as a method for user authentication has been critically investigated, translated into design solutions, and evaluated in a controlled, longitudinal study.

2. Visual memory

The VIP idea has arisen from the knowledge that visual memory is extremely powerful. Classic cognitive science studies have shown that humans have a vast, almost limitless memory for pictures in particular. Pictures are usually remembered far better than words [9, 10, 12, 13], and visual memory does not seems to be significantly affected by the general decline of cognitive capabilities associated with ageing as occurs with other types of memory [11].

The memory system in the human brain can be regarded as 3 basic stages. Firstly there is an encoding stage, where a memory is laid down. This is the learning stage and it is possible to have some influence on it by changing the way in which a given stimulus is presented. The next stage is the storage of information and it is difficult to have influence on this stage. Retrieval is the third and last stage. Memory is retrieved from the brain through association and this process is the most open to improvement. The basic two retrieval processes are recall and recognition (the awareness that an object or event is one that has been previously seen, experienced or learned).

The three memory stages differ according to the nature of the information to be processed. Alphanumerical symbols are encoded, stored and retrieved differently than pictures. The superiority of pictures over words has been attributed to encoding differences between the two symbolic formats [10, 13]. Pictures engage greater conceptual elaborative processing than words and explicit retrieval is enhanced under these conditions. Pictures may engage greater elaboration because they are associated with more symbolic codes or with a more distinctive code than words. A difference in the way words and pictures are stored the brain has also been hypothesised. Finally, it is well known that pictures and words are better retrieved under different conditions: free recall suits alphanumeric stimuli, recognition suits pictures.

The idea of using images as substitutes for PINs raises a number of issues. Some of them relate to the security of the transaction, others refer to cognitive constraints affecting memory, and to the acceptability of the solution. Research in cognitive psychology provides a broad understanding of how memory works, but this knowledge did not prove to be enough to inform the design of the visual PIN. Indeed, current knowledge derives from controlled experiments, which are difficult to apply to real life situations. In particular, most of the research is based on a two-alternative forced-choice paradigm. Participants are first required to learn a number of items displayed individually. Then, during the memory test, they are shown test pairs of stimuli, consisting of a new and an old stimulus. At this point, they have to indicate which of the stimuli they had previously seen and which they had not seen before. It is clear that this way of presenting information cannot be applied to a self-service situation where execution speed is paramount.

More studies are required to understand memory performance when the target images have to be recognised among a set of distractors, as required by the PIN paradigm. Such an effort appears worthwhile if one considers the security advantages of the solution which add to the expected mnemonic improvement. Pictures are often difficult to describe verbally, which should discourage people from revealing or writing down their code

3. Graphical password

The idea of using graphical passwords is not new. In 1996 Blonder patented a graphical password which requires the user to touch predetermined areas of an image in a fixed sequence for authentication [8]. Jeremy and colleagues implemented the concept on a PDA, so exploiting the input capabilities of graphical devices. The password consisted of a simple picture drawn on a screen [8].

Passlogix[1] Inc. distributes V-go, an application that allows users to create passwords simulating familiar actions while clicking on objects of graphical interfaces. For example, users can mix a cocktail, cook a meal or dial a phone number: the password corresponds to the sequence of the objects they had clicked on. Passfaces by IDArts[2] is based on the face recognition and is currently freely available on the Internet. Users are given 'five faces', which they have to recognise amongst a set of distractors. Each 'face' of the password is presented on a separate screen amongst different distractors, so that sequence retrieval is controlled by the system. This implementation however may be time consuming and does not suit self-service environments.

Despite the examples of graphical passwords already on the market are claiming exceptional reliability and ease of use, very few user studies are available to explain how people actually use these systems. Dhamija and colleagues [5] have investigated the memorability of pictures against passwords and PINs. Two types of images were tested: abstract and photographic pictures. The experiment involved 20 participants and consisted of two sessions. In session one, participants had to create a password of

[1] http://www.passlogix.com

[2] http://www.realuser.com/cgi-bin/pcenter.exe/_/index.htm

at least six characters and a four-digit PIN, both of which were believed to be secure and never used before. Participants also selected two image portfolios of 5 pictures each. One was picked up from a set of one hundred abstract images and the other from one hundred photographic pictures. Participants finally had to authenticate using all four techniques. In the portfolio conditions, they had to select their images from a challenge set of 25 pictures. One week later, participants had to log in again using all four techniques.

Results of this study showed that creating passwords and PINs is much faster than creating an image portfolio, with the photographic conditions requiring the longest time. The mnemonic advantage of visual recognition became evident after the week interval: 7 participants failed to recall the password, 6 the PIN, 2 the abstract image portfolio and 1 the photographic portfolio.

The study is interesting but it is clearly more focused on security issues than on usability. In particular, the research method does not seem completely satisfactory. We are concerned with the use of a within-subject design in which each participant had to remember simultaneously 4 different codes, a situation that may lead to uncontrolled mnemonic interference. Moreover, the procedure is not fully explained. It is not clear, for example, if (and eventually how) participants were induced to memorise their codes and if any learning session took place. In our view, the study addressed implicit memory of symbolic and visual material. Finally, the focus of the paper is on qualitative findings and subjective interpretations. The lack of statistical information makes it difficult to understand the strength and reliability of the reported findings.

Emerging alternatives to the PIN approach based on visual memory warrant further user-centred investigation in order to address user satisfaction and performance. There are still several questions about the use of visual passwords left unanswered. Some of them have been addressed in our experiment.

4. The VIP approach

VIP is an innovative concept for user authentication, based on the psychological assumption that pictorial recognition is easier than the recall of numbers or passwords. The user is given an image portfolio, which represents their password. To authenticate, the user must correctly identify the images that are part of their portfolio inside a wider challenge set randomly selected from a visual database. The most innovative features of the design are described and justified below.

- The pictures are detailed, colourful and meaningful photos of objects.

This contrasts with the claim of Dhamija and colleagues [5] that abstract images are more secure. In their view, realistic pictures may be more easily communicated to others than abstract ones. Photographic pictures may also induce a biased selection reflecting personal preferences, which can be easier to predict. Our choice is justified by the fact that most of the studies in the memory literature have used colourful, detailed pictures of real objects for best performance. Users also appear to like photographic pictures better [5]. The predictability issue, which however still requires empirical validation, has been resolved by giving the users a randomly generated portfolio.

- Images stored in the visual database are clustered in semantic categories (e.g., flowers, animals, rocks, landscapes etc).

This helps controlling the visual configuration of the challenge set. In particular, it allows establishing precise rules to determine the ratio between targets (picture which form the visual code) and distractors randomly selected from the database.

- The code is composed of a sequence of pictures.

Having to recognise a number of objects in a sequence adds a further level of security to the transaction. The capability of people handling such a constraint was one of our experimental questions.

- Users are assigned their code.

We believe that the role of motivational factors related to the fact of choosing ‘personal images’ should not seriously affect performance. Indeed, the picture superiority effect is very robust even when people are required to learn hundreds of pictures [12]. Further, this choice greatly improves the security of the transaction by avoiding the selection of codes that are too simplistic and may be guessed by others. Finally, it strongly reduces the time needed for code selection, which may affect system efficiency [5]. An automatic enrolment procedure was proposed: customers were given their code directly at the ATM screen without the need for printouts.

- The code is entered using a touch-screen interface.

The use of touch-screens as entry devices for PIN is still controversial and does not comply with current international standards for ATM transactions. Furthermore, from a security point of view, PIN-entry devices should be as close to horizontal as possible, to prevent shoulder-surfers from spying the code. Nevertheless, we believe that touch-screen is the ideal solution for the VIP paradigm, especially if the visual code is composed of a randomly positioned and selected image set. In real life environments, the security of the transactions may be enhanced by using privacy filters and by consumer education.

5. User evaluation

The project has investigated attitudinal, cognitive, usability and security issues related to the VIP approach in comparison with the traditional PIN. The experiment was designed to answer the following questions.

Q1. Are visual images more effective than numbers as a means for user identification?

Q2. Is VIP more secure or perceived as more secure than traditional PIN?

Q3. Do people prefer VIP to PIN?

Q4. What is the effect of location on image recognition?

Q5. Can people retrieve a sequence of pictures?

Question 4 refers to the role of motor memory and of memory for spatial location [6]. The act of entering a code involves more than number recall or pictures recognition. It requires two basic actions that may be performed in parallel: the code is retrieved and is entered on the keypad. Without realising it, people use an implicit memory related to the movement performed and to the position of the objects.

Four different systems/implementations were designed and compared. They correspond to experimental conditions and are summarised in Table 1.

Table 1. Systems tested in the evaluation

System	Type of code	Location
PIN	4 fixed order numbers from 10	Fixed
VIP1	4 fixed order images from 10	Fixed
VIP2	4 fixed order images from 10	Random
VIP3	Portfolio based	Random

The first condition (PIN) is a touch screen implementation of the traditional PIN systems. Participants are asked to learn a set of 4 digits and recall them in order. The digits are displayed in the same layout as telephone or ATM number-pads (Figure 1).

Figure 1. The PIN interface

All the other conditions substitute numbers with pictures. VIP1 is the pictorial equivalent of the PIN paradigm (Figure 2). It requires the user to memorise a sequence of 4 pictures, which are always displayed in the same location of the visual keypad and must be entered in a fixed order (Figure 2). The interface resembles the PIN keypad but a new set of distractors is randomly extracted from the visual database whenever the user makes an authentication attempt (Figure 2). To minimise mnemonic interference, each picture of the authentication code belongs to a different semantic category and the distractors are selected from the remaining categories.

VIP2 differs from VIP1 in that the 4 pictures forming the authentication code are displayed in new random positions around the set of 10 locations of the visual keypad at each authentication attempt (Figure 2).

Figure 2. The challenge set in condition VIP1 and VIP2

VIP3 is a different concept, which was designed to investigate the limits of the visual paradigm. The user is assigned a portfolio of 8 pictures. At every authentication attempt, four of these pictures are randomly displayed together with 12 distractors in the challenge set. The distractors are randomly selected from the database, avoiding however the categories of the targets currently displayed in the challenge set. To authenticate, the users have to select their images from the 16 shown on the interface, in any order. A screen shot of the challenge set is reported in Figure 3.

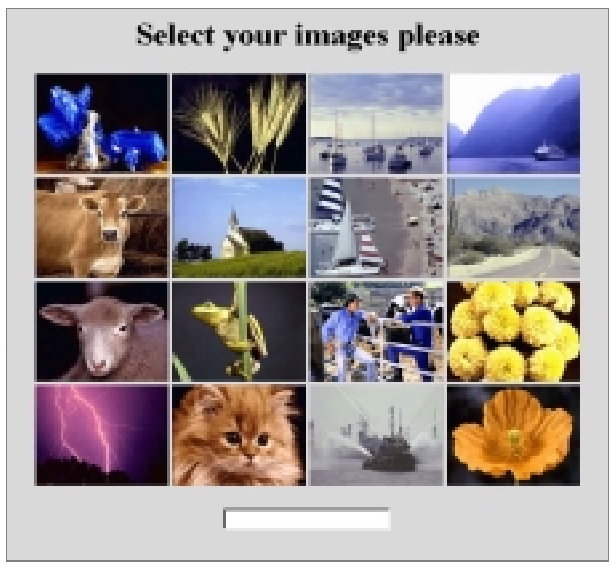

Figure 3.The challenge set in condition VIP3

The comparison between PIN and VIP1 allowed us to answer the first three experimental questions. We expected pictures to be a viable alternative to numbers: they should be easier to remember and provide a more enjoyable user experience (*Picture superiority hypothesis*).

The comparison between VIP1 and VIP2 allowed testing of the effect of location on code retrieval (Q4). We believed that both motor memory and memory for locations may implicitly facilitate the retrieval of a sequence of targets. Hence, we expected an advantage of VIP1, fixed location, over VIP2, random location, (*Fixed location hypothesis*).

The comparison between VIP2 and VIP3 tested the conjoint effects of visual code size, sequence and challenge set size.

5.1 Procedure

The experiment involved 61 participants who attended two sessions separated by a week interval. Participants were recruited by a brief phone interview to guarantee that all of them were ATM users and did not have pathological memory deficits. The first session lasted almost an hour. After filling in a questionnaire describing their general behaviour and attitude towards ATMs and PINs, participants were randomly assigned to one of the four experimental conditions and given an ATM card. Swiping it, they first underwent an automatic enrolment session, designed to help them learn their code.

After enrolment, participants performed 10 authentication trials (learning phase). They had to swipe their card and enter the code as fast and accurately as possible. At the end, a questionnaire collected their initial impressions of the system.

The first memory test (test 1) took place 40 minutes later, after participants had performed a distractor task, namely interacting with a chatterbot, a computer program that simulates a typed conversation with the user. The second memory test (test 2) took place a week later and was followed by a questionnaire to assess their final opinions. During both test sessions, participants were invited to authenticate, swiping their card and entering their code 10 times in a row, as fast and accurately as possible. As in a normal PIN transaction, in case of erroneous code selections, participants were automatically given up to 3 attempts.

Behavioural data (errors, reaction time and entry time) were collected by automatic logging. Subjective evaluations were collected by a battery of psychometric instruments developed for this research.

5.2 Design

The comparative evaluation of the systems was based on a between-subjects design. Participants were randomly assigned to one of the four experimental conditions previously described (PIN, VIP1, VIP2, VIP3). Data were collected at three stages: learning, test1 and test2.

5.3 Results

Participants were 29 males and 32 females, covering a broad range of ages (from 16 to 66 years, mean = 30), and education levels. All participants were ATM users and 74% used the ATM at least once a week. They reported to use an average of 4 different PINs or passwords for a variety of devices, such as mobile phones, computers and ATMs.

PIN usage in the ATM context appeared to be somehow problematic. Some 37% of the participants have had their card retained by an ATM because they were unable to remember the correct PIN. The main reasons were a mismatch between PIN and card, inexperience or very occasional use. Furthermore, 50% of the sample declared they had difficulty remembering their bank PIN and 36% admitted to having communicated their PIN to another person.

The evaluation metrics was defined along the three major dimensions defining usability: *effectiveness*, *efficiency*, and *user satisfaction*. Effectiveness is here associated with code memorability and defined in terms of number of people who forgot their security code and numbers of wrong entries. Efficiency refers to speed of data entry. User satisfaction refers to the perception of the system relative to the perception of traditional keypad based PIN devices.

5.3.1 Effectiveness

In contrast to [5], none of the participants in our experiment ever forgot their authentication code. The difference can be attributed to the intensive training session of our study, which aimed to reproduce a condition of frequent use in a short period of time.

However, the performance was not entirely error free. In almost 5% of the authentication trials (118/2196) the users could not enter the correct code. A crosstabulation analysis indicated that these errors were not homogeneously distributed among the four conditions, $\chi^2_{(3)} = 57.08$, $p < .001$. Rather, they tended to concentrate in condition VIP3, which accounted for more errors than all the other three conditions together (Figure 4). A slight advantage of VIP1 over VIP2 was also observed, $\chi^2_{(1)} = 3.42$, $p = .07$. It supports the fixed location hypothesis. Contrary to our expectations, no differences between numbers and pictures (PIN vs. VIP1) were observed.

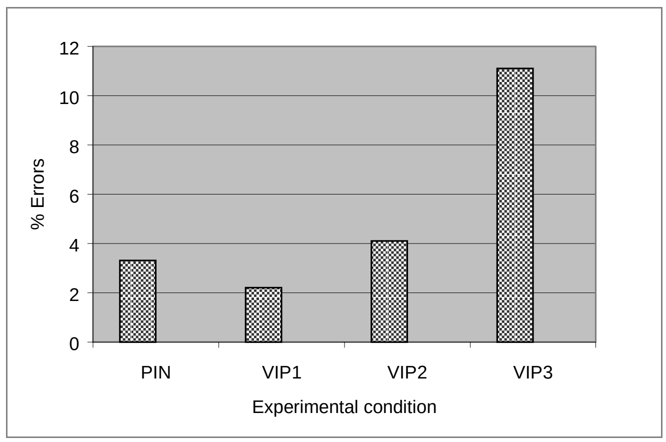

Figure 4. Percentage of errors in the experimental conditions

A detailed error analysis evinced the occurrence of four basic error types. They are described below and their distribution is illustrated in Figure 5.

- *Sequence*: the correct code is retrieved but entered in a wrong order.
- *Double click*: the same item is unintentionally selected two consecutive times (the prototype did not allow corrections).
- *Double selection*: the same item is selected twice in non-consecutive positions.
- *Wrong selection*: one or more of the selected items do not belong to the authentication code.

Analysing the graph in Figure 5 it appears that different system configurations trigger specific error types. The poor performance of the VIP3 condition was mainly due to wrong selections. In particular, people tended to falsely recognise distractors belonging to the same category of items of their code, which were not displayed in the current challenge set. This finding indicate that visual memory is sensitive to interference so that if participants had a flower in their portfolio, they were induced to identify other flowers in the challenging set as 'their flower' Inter-category wrong-selections also occurred, particularly when the targets were not entirely meaningful (e.g., rocks and minerals). In this case targets were confused with distractors which had very similar visual configurations even if they belonged to other semantic categories (e.g., a yellow flower mistaken for a yellow mineral).

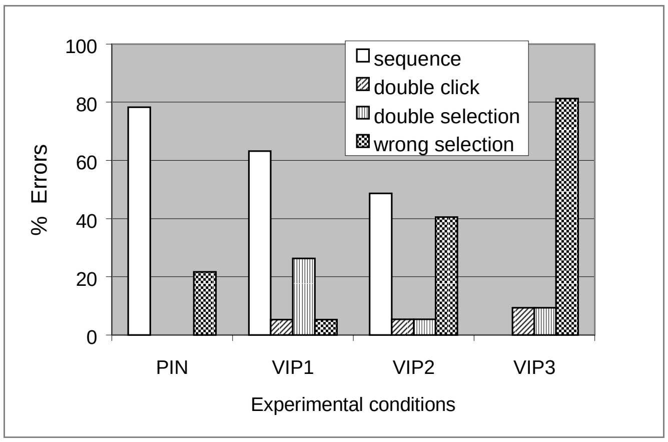

Figure 5. Types of Errors

The distribution of errors as a function of experimental stages (learning, test1 and test2) is illustrated in Figure 6. Wrong actions tended to occur at different moments in the four experimental conditions, $\chi^2_{(6)}= 29.10$, $p <.001$..

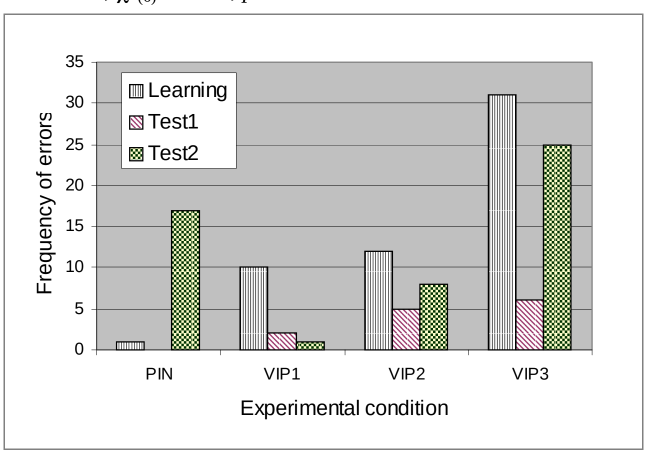

Figure 6. Error occurrence at different experimental stages

PIN and VIP1 presented an interesting difference in error distribution. In the PIN condition, most of the errors occurred after the week interval, when some participants struggled to retrieve their numeric code. In the VIP1 condition, most of the errors occurred during the learning session, when participants had to familiarise themselves with the innovative systems. Thereafter, errors tended to disappear, with only one occurrence after a week. This result supports the picture superiority hypothesis.

5.3.2 Efficiency

Two variables were analysed to evaluate the efficiency of the authentication methods:

- Reaction time: lag between the appearance of the challenge set on the screen and the selection of the first item of the code.
- Entry time: lag between the first and the last selection.

For every participant, these variables were computed averaging the timing of correct actions in each experimental phase. They were then entered as dependent variables in a separate mixed design ANOVA, with experimental stage (3) as the within-subjects factor and system (4) as the between-subjects factor. Post-hoc analyses based on the LSD model (Least Significance Difference) were also performed to test specific hypotheses (PIN vs. VIP1 pictures superiority hyp.; VIP1 vs. VIP2 fixed location hyp.; VIP2 vs. VIP3 limit of memory).

As regards reaction time, both the main effects were highly significant, namely stage $F_{(2,114)}= 9.22$, $p < .001$; system $F_{(3,57)} = 5.73$, $p <.01$. The 2-way interaction was a tendency, $F_{(6,114)} = 1.73$, $p =.12$. The effect of stage indicates that reaction time changed during the experiment. The system effect indicates that the design solution influenced the time needed to locate and select the first item of the code. The weak interaction suggests that the main effects may influence each other: different designs solutions affected reaction times differently at different stages.

Figure 7 displays the average reaction time as a function of system and stage. In every condition, participants tended to be much slower during the learning phase than during test1. Participants in the VIP1 and VIP2 condition remained stable after the week interval, while participants in the PIN and VIP3 condition tended to slow down suggesting a stronger effort in retrieving the code.

Participants who used the VIP3 system constantly achieved the slowest performance. The effect is presumably related to the need for visual scanning of the challenge set in order to locate an item, without knowing what items to look for.

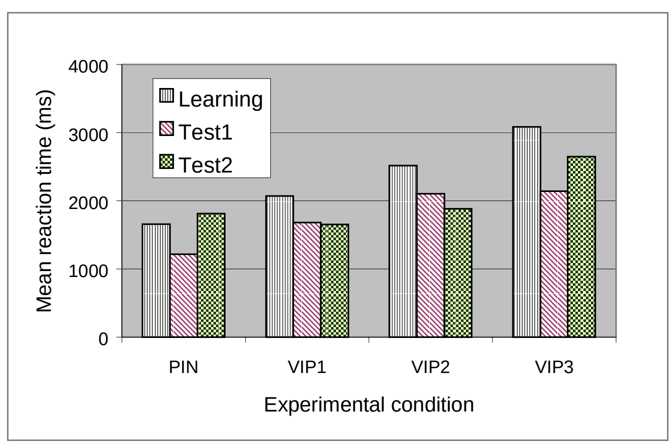

Figure 7. Reaction time as a function of stage and system

Post-hoc comparisons showed no significant differences in reaction time between PIN and VIP1 or VIP1 and VIP2. Contrary to our expectations, reaction time did not appear to be influenced neither by the nature of the item to be located nor by the knowledge of its position.

A different pattern emerged by the post-hoc analyses run on entry time. In this case, participants in condition VIP1 were significantly faster than participants in condition VIP2 ($p < .05$). This finding confirmed that fixed location speeds up performance. No difference emerged between users who had to remember pictures and users who had to remember numbers (PIN vs. VIP1).

A significant effect of stage [$F_{(2,114)} = 15.35$, $p < .001$] and system [$F_{(3,57)} = 25.26$, $p < .001$] emerged also from the Anova on entry time. The 2- way interaction was not significant. The effect of stage was clearly due to the improvement from the learning to the test phases. The effect of system was due to the particularly slow performance of participants in the VIP3 condition (Figure 8). Yet again, VIP3 was penalised by the need for visual scanning over a larger challenge set and by the lack of knowledge of what items were displayed.

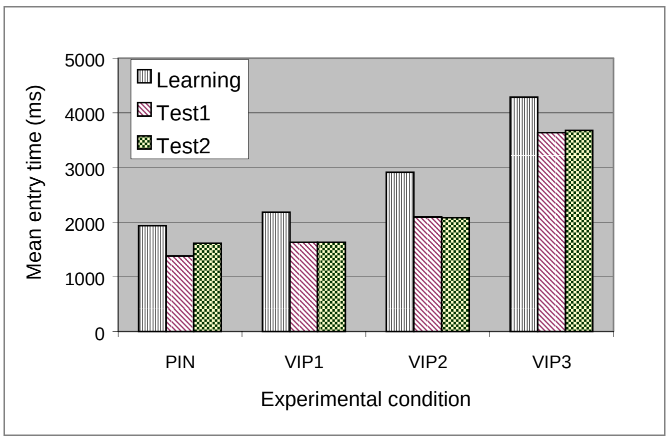

Figure 8. Entry times as a function of stage and system

Analysing the global picture of performance data related to effectiveness and efficiency, one can notice that the performance was not affected by speed-accuracy trade-off. Whenever the time increased there were also more error (Figure, 6; Figure 7; Figure 8).

5.3.3 User satisfaction

For each user, two basic measures were assessed:

- Satisfaction with the traditional PIN (i.e., number-pad implementation of current ATM users had experienced before the experiment).
- Satisfaction with the device tested during the experiment (i.e., one of the four experimental systems).

Opinions and attitudes towards the traditional PIN were used as baseline values. This procedure allows measuring perceived advantages or disadvantages of the tested solution relative to the current one.

Attitudes were measured by 7 items of a semantic differential scale covering a number of usability and security dimensions[3]. The reliability of the scale was high in both the administrations ($\alpha > .80$) with all items presenting a satisfactory item-scale correlation index. Therefore, two attitude indexes were calculated averaging ratings for the individual items. The index ranged from 1 (extremely negative attitude) to 7 (extremely positive). The mean values are displayed in Figure 9. Note that all the evaluations are in the positive half of the scale, reflecting positive attitudes towards the targets.

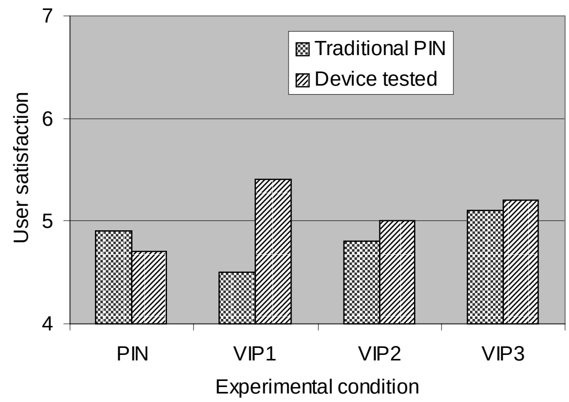

Figure 9. User satisfaction

To compare the user satisfaction in the 4 experimental conditions, the difference between attitudes towards traditional PIN and attitudes towards the new system was calculated. The variable was then analysed by an ANOVA with system (4) as the between-subjects factor. Results indicated a marginally significant effect of System, $F_{(3,59)} = 2.58$, $p = .06$. As can be seen in Figure 9, this effect is due to the strong improvement of attitudes in participants who used the VIP1 system.

Further analyses based on different measurement techniques demonstrated that participants perceived the visual code as easier to remember, more secure and in general preferred over the numeric code. The advantage of VIP1 over other visual configurations was constant. Sequence retrieval in pictorial configuration was not perceived as a problem. According to follow-up interviews, users developed an easy strategy to support it: individual pictures were incorporated into a narrative.

6. Conclusion

This study and associated design exploration has revealed important information about consumers and their attitudes towards and behaviour with PINs, including novel approaches using images. The evaluation provided detailed data on the memorability of images and a deep insight on cognitive constraints of visual and numerical memory in the context of self-service.

Pictures tended to be less error prone than numbers after a week interval and did not compromise the speed of the transaction. No

[3] The following couples of adjectives were used: secure–insecure; difficult–easy; satisfactory–unsatisfactory; slow–fast; boring– fun; relaxing–stressful; inefficient-efficient.

difficulties emerged with respect to sequence retrieval in visual code recognition. The user reaction to the VIP concept was promising. Overall, users liked the VIP concept better and declared that it was more secure and easy to remember than the PIN. Although these reactions need to be weighted taking into account the novelty factors, they suggest widespread acceptance of the VIP paradigm.

The comparative evaluation of different implementations of the VIP concept provided a number of insights, which may help to make the best use of the natural power of visual memory in security solutions. The 'worst' design condition as regards performance criteria was VIP3: the user had to remember a portfolio of 8 pictures without any sequence. Four of them were randomly displayed together with 12 distractors. Two basic factors can be held responsible for such a poor performance. Firstly, visual memory is very sensitive to interference so that if participants have a flower in their portfolio, they tended to identify each flower in the challenging set as 'their flower'. Secondly, in this condition no learning is possible, since people do not know which part of their code will be displayed and where the targets will appear. Therefore, they need to scan all the visual display before planning their action.

The interference effect can be explained considering that memory is a constructive process, meaning that details that are not held in memory can be added later to that memory. For instance, if we are asked to remember the picture of a room, our memory will not hold all the details of that room. It will, however, remember the important details and piece together the rest of the details from what makes sense. The design implication of this finding supports the choice of detailed, colourful, and meaningful photos of real objects. It also suggests categorising the visual database to avoid displaying interference prone images on the challenge set.

The 'best' visual condition as regards objective and subjective evaluation is VIP1: the subject had to retrieve 4 pictures out of 10 in the correct sequence. Each picture is always displayed in the same position. The positive effect of fixed location on performance and subjective evaluations should not be under-estimated. Fixed location decreases the amount of errors and speeds up the action of entering the code. The performance benefit may be due to: (a) memory for spatial location; (b) motor memory; or a (c) combination of both factors. Further research is needed to better understand this effect.

To conclude, our data demonstrated the merit of the visual PIN idea but stressed the need for further user-centred design to maximise the benefit of the concept. Indeed, we have demonstrated that the benefits of using pictures instead of numbers may be easily disrupted by a wrong design, as in the case of the VIP3 systems.

7. References

[1] Adams, D.A. and Chang, S.Y. An investigation of keypad interface security. *Information & Management* 24 (1993), 53-59.

[2] Adams, A. and Sasse, M.A. Users are not the enemy. *Commun. ACM* 42 (December 1999), 41-46.

[3] Ashbourn, J. *Biometrics. Advanced Identity Verification*. Springer Verlag, London, 2000.

[4] Baddeley, A.D. *Working memory*. Oxford University Press, 1990.

[5] Dhamija, R. & Perrig, A. Déjà vu: A User Study Using Images for Authentication. In *Proceedings of 9th USENIX Security Symposium*, August 2000.

[6] Fisher, M. H. Probing spatial working memory with Corsi Blocks task. *Brain and Cognition* 4 2001, 143-154.

[7] Gong, L., Lomas, M.A., Needham, R.M. and Saltzer, J.H. Protecting poorly chosen secrets from guessing attacks. *IEEE J. on Selected Areas in Communications*, 11(5), 1993, 648 - 656.

[8] Jermyn, I., Mayer, A., Monrose, F., Reiter, M.K., & Rubin, A.D. The design and Analysis of Graphical Passwords. *Proceedings of the 9th USENIX Security Symposium*, August 2000.

[9] Madigan, S. Picture memory. In J.C. Yuille (Ed.), *Imagery, memory, and cognition: essays in honor of Allan Paivio*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1983.

[10] Paivio, A, Rogers, T.B., & Smythe, P.C. Why are pictures easier to recall than words? *Psychonomic Science*, 11(4), 1968, 137-138.

[11] Park, D.C. Ageing and memory: Mechanisms underlying age differences in performances. In *Proceedings of the 1997 World Congress of Gerontology*.

[12] Shepard, R.N. Recognition memory for words, sentences and pictures. *Journal of Verbal Learning and Verbal Behavior*, 6, 1967, 156-163.

[13] Vaidja, C.J. and Gabrieli, J.D. Picture superiority in conceptual memory: Dissociative effects of encoding and retrieval tasks. *Memory and Cognition*, 28(7), 2000, 1165-1172.

[14] Yan, J., Blackwell, A., Anderson, R. and Grant, A. The memorability and security of passwords – Some empirical results. *Technical Report No. 500 2001*, Computer Laboratory University of Cambridge, http://www.ftp.cl.cam.ac.uk/ftp/users/rja14/tr500.pdf.