Application of the Technology Acceptance Model to OCR-based CAPTCHA systems

Pawel Lupkowski\textsuperscript{a,}\textsuperscript{*}, Pawel Kleka\textsuperscript{a}

\textsuperscript{a}Institute of Psychology, Adam Mickiewicz University, Szamarzewskiego 89a, 60-568 Poznan, Poland

Abstract
This research investigates the applicability of F. Davis’s Technology Acceptance Model to OCR-based CAPTCHAs involved in free e-mail accounts registration process. The results provide preliminary evidence that Technology Acceptance Model may be a useful tool for researches concentrated on CAPTCHA’s acceptance. Results are especially interesting when it comes to perceived usefulness concept of TAM and its direct relationship with attitude toward use concept. This strong relation indicates a new area of research for CAPTCHA developers interested in improving usability and accessibility characteristics of those systems.

Keywords: CAPTCHA, Turing test, Technology Acceptance Model, usability, accessibility

1. Introduction

1.1. Aim of the study
The aim of this study is to investigate the applicability of Fred Davis’ Technology Acceptance Model to OCR-based CAPTCHAs involved in free e-mail account registration process. In our opinion Technology Acceptance Model research might bring new and fruitful ideas into CAPTCHA’s development area, especially when it comes to user-friendliness aspects of CAPTCHA systems.

\textsuperscript{*}Corresponding author
Email addresses: Pawel.Lupkowski@amu.edu.pl (Pawel Lupkowski), Pawel.Kleka@amu.edu.pl (Pawel Kleka)

Preprint submitted to Elsevier October 28, 2010
1.2. What is a CAPTCHA?

CAPTCHA stands for Completely Automated Turing Test To Tell Computers and Humans Apart. The very idea of CAPTCHA systems comes from the seminal paper of Moni Naor Verification of a human in the loop or Identification via The Turing Test [18], while the name was coined by Luis von Ahn. The main task of a CAPTCHA is to automatically differentiate bots (malicious programs) and human users in on-line services. There are many domains where such systems are needed, like for example, commenting blogs, Internet messages boards, sending SMS/MMS messages via web-pages, on-line polls etc. However the most important domain, where providers want to be sure that they are coping with human users are free e-mail accounts. CAPTCHAs are widely used by free e-mail accounts providers to stop SPAM messages being sent from fictional accounts registered by bots. Users of Gmail, Yahoo! and many other popular providers in this field have to solve CAPTCHA task while registering e-mail account.

The most characteristic properties of CAPTCHA are (cf. [1], [2], [18]):

- test instances and their solutions are generated automatically;
- most humans can pass the test in a very short period;
- it is difficult to write a computer program that will be successful in passing the test in question;
- the underlying problem must be difficult for programs used currently as well as those which will be developed in the future.

Nowadays, only one CAPTCHA scheme become popular, namely so called OCR-based CAPTCHA (often referred to as ‘visual CAPTCHAs’). Tasks of this kind are based on identifying the content of a distorted picture, like recognizing a text (most popular, used e.g. by Yahoo! or Hotmail) a common animal (ANIMAL-PIX, a CAPTCHA which can be found at www.captcha.net), or a geometric figure. This kind of CAPTCHA tasks exploits the ability of people to read images of text more reliably than optical character recognition (OCR) or other machine vision systems (cf. [7]). Example of this kind of CAPTCHA (used by Yahoo!) is presented in Figure 1. It is worth to stress that CAPTCHA become popular not only on large portals but it is also a kind of standard anti-SPAM solution on numerous of small (even private) web-sites.
Unfortunately, the ability exploited by visual CAPTCHAs seems not to be reliable enough any more. There is a significant number of papers presenting attacks on well known visual CAPTCHAs schemes. Let us mention only some of them. Mori and Malik [17] describe an attack on the EZ-Gimpy CAPTCHA used by Yahoo!, which enjoyed a success rate of 92%. In the more difficult case of Gimpy they passed the test 33% of the time. Moy et al [16] achieved a 99% rate of solved tasks of EZ-Gimpy (and on Gimpy-r — a four letter version of GIMPY — 78%). Yan and Ahmad in [24] present a segmentation method that enabled them to break the Microsoft CAPTCHA scheme with an overall success rate of 60%. There are works like [23] where quite simple techniques are used to break many visual CAPTCHAs. This leads to a situation where CAPTCHA tasks become more and more complicated (with motivation to assure more security). As a result many of tasks offered to users are not even human-readable. To give a simple example, we’ve performed a simple research. 10 participants had to solve 10 CAPTCHA tasks each (retrieved from one of the biggest Polish e-mail accounts providers). Only correctness of solution was measured (time was unlimited). Tasks were presented on a web page and participants solved them using their home computers. In a result only 3 tasks were solved correctly.

As a consequence of the growing popularity combined with growing difficulty of CAPTCHAs, problems of accessibility and usability of this technology becomes more and more important. There are papers describing research in this field, where some design rules are discussed to make CAPTCHAs more accessible (cf. [13]) and more usable ([25], [5], [6], [12], [22]).

What is more, since CAPTCHAs are obligatory elements of registration process, issues like user’s attitudes, feelings and acceptance of this kind of tasks also should become a relevant element of CAPTCHA design and implementation in a given service. In our opinion there is a need for a tool that will concentrate on user’s attitudes and a level of acceptance for CAPTCHA
systems. What is more this tool should also allow for identifying factors responsible for users’ attitudes. Davis’ technology acceptance model seems to be a good candidate for such a tool.

1.3. Technology Acceptance Model

Technology acceptance model (TAM) was proposed by Fred Davis in [9]. The main aim of this model is to provide a tool for assessing Information Technology (IT) acceptance among users. The model (as it is described in [10]) is presented in Figure 2.

It consists of the following components: External Variables (system and/or users characteristics), Perceived Usefulness, Perceived Ease of Use, Attitude Towards Using, Behavioral Intention to Use, and Actual System Use. *Perceived Usefulness* (PU) is defined as “the degree to which person believes that using a particular system would enhance his or her job performance” [9, p. 320]. *Perceived Ease of Use* (PEoU) is understood as “the degree to which a person believes that using a particular system would be free of effort” [9, p. 320]. *Attitude Towards Using* (ATU) is “the degree of evaluative effect that an individual associates with using the target system in his or her job” [11, p. 476]. Davis developed reliable scales for the model’s components (cf. [27]).

TAM is widely used to investigate and predict users’ acceptance for various IT systems, like text editors and e-mail clients [9], knowledge management systems [15], multimedia learning environments [20], FireFox Internet browser [21], or wireless technology [26]. Also many extensions and revisions of TAM are proposed (see for example [27], [14], [19], [8]).
2. Research

2.1. Model and hypotheses

For the research we have decided to use the technology acceptance model in its form presented in the previous section. We have resigned from the ‘External variables’ module. The reason for this is that our research was focused on the class of CAPTCHA systems (namely OCR-based systems involved in an e-mail account registration process) not on a given piece of software (like in original Davis’ research).

Our research hypotheses — drawn from previous TAM research — were:

- (H1) Positive relationships will be observed between PU and BI
- (H2) Positive relationships will be observed between PEoU and ATU
- (H3) Positive relationships will be observed between PU and ATU
- (H4) Positive relationships will be observed between ATU and BI
- (H5) Positive relationships will be observed between BI and AU
- (H6) Positive relationships will be observed between PEoU and PU

2.2. Development of the scale

21 item survey was prepared for the research. Items and measurement scales were based on the previous TAM research ([11], [21], [27], [14]) and tailored to the CAPTCHA’s specific area. For Perceived Usefulness (PU), Perceived Ease of Use (PEoU) and Attitude Towards Using (ATU) four item scales were used. Behavioral Intentions to Use (BI) and Actual System Use (AU) were measured by three item scales. Four items concerned questions about participant (sex, age, number of used e-mail account, self esteemed computer expertise). We have used two versions of the on-line questionnaire. After obtaining first results it appeared that some items should be paraphrased (for details see Section 3.1), after that, the improved version of the questionnaire was published. The TAM scales of the final version of the questionnaire are shown in Appendix.
Table 1: Self esteemed level of expertise in a computer usage (group one)

<table>
<thead>
<tr>
<th>Level</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (beginner)</td>
<td>0%</td>
</tr>
<tr>
<td>2</td>
<td>2%</td>
</tr>
<tr>
<td>3</td>
<td>29%</td>
</tr>
<tr>
<td>4</td>
<td>43%</td>
</tr>
<tr>
<td>5 (advanced)</td>
<td>27%</td>
</tr>
</tbody>
</table>

2.2.1. Procedure

The questionnaire was powered by GoogleDocs form and put on-line with additional explanations about OCR-CAPTCHAs within a web-page prepared for the research (published on the Adam Mickiewicz University server). Link for the page was sent to popular websites (like facebook) and to portals and forums devoted to IT, psychology and cognitive science, where CAPTCHA issues were discussed. The questionnaire was accessible for 16 weeks. The web-page was visited 259 times. 165 valid responses were collected.

2.2.2. Participants

Responses were obtained from 163 participants. 112 for the first version of the questionnaire — 50 men, 62 women, aged form 16 to 60 (60% of the group had 20-24 years, average age amount $M = 24.7, SD = 6.5$). Hereafter we will refer to those participants as a group one. After the scale validation a modified version of questionnaire was published on-line. At this stage responses were obtained from 51 participants — 21 men, 30 women, average age within this group was 23.6 years ($SD = 3.79$). This group of participants will be referred to as group two.

Participants were asked to estimate their self esteemed computer expertise. Detailed results (for the first group) are presented in Table 1 (the scale was 1 — beginner to 5 — expert).

Results for the second group of participants is presented in Table 2 (here, after modification of questionnaire, scale was 1 — beginner to 7 — expert).

The research was focused on CAPTCHAs connected to the free e-mail accounts providers, thus we have also asked how many e-mail accounts does a subject have. Average result was 3.1 (from 1–10 e-mail accounts). Men characterized in higher ($t_{156} = 4.2, p < 0.001$) self esteemed expertise in computer usage ($M_F = 3.9, SD_F = 1.02; M_M = 4.6, SD_M = 0.93$) than women. There were no significant relationship between a level of self esteemed level
Table 2: Self esteemed level of expertise in a computer usage (group two)

<table>
<thead>
<tr>
<th></th>
<th>N=51</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (beginner)</td>
<td>0%</td>
</tr>
<tr>
<td>2</td>
<td>0%</td>
</tr>
<tr>
<td>3</td>
<td>16%</td>
</tr>
<tr>
<td>4</td>
<td>25%</td>
</tr>
<tr>
<td>5</td>
<td>27%</td>
</tr>
<tr>
<td>6</td>
<td>22%</td>
</tr>
<tr>
<td>7 (advanced)</td>
<td>10%</td>
</tr>
</tbody>
</table>

of expertise in computer usage and a number of e-mail accounts in a group of men, and very weak relationship in a woman group (Spearman’s $\rho = 0.30$, $p < 0.01$). The were non significant differences between groups except of the self esteemed level of expertise in a computer usage — higher results were observed in the second group ($M_{g1} = 4.8$, $SD_{g1} = 1.22$ vs $M_{g2} = 3.9$, $SD_{g2} = 0.8$, $t_{70} = 4.80$, $p < 0.05$).

We may say that our average subject was a young person with high level of computer expertise, actively using e-mail account, aware of what CAPTCHA is and with experience of solving CAPTCHA tasks.

3. Results and discussion

3.1. Scale validation

After collecting the data from the first version of the questionnaire it appeared that it had low reliability for some of questions — cf. Table 3. To improve this results we have decided to rewrite negatively phrased items of the questionnaire (which is suggested by F. Davis in [9, p. 327]). Analysis of the data for the improved questionnaire showed much improvement of reliability for questionnaire questions, with the whole scale reliability improvement from 0.87 to 0.90 — cf. Tab. 3. Only exception here was the Actual Use construct, which reliability was very low for the first and for the second version of the questionnaire. The reliability was slightly improved by one item removal, but it still remained low (in comparison to other constructs). This might be the effect of a different research scenario used. Typically the TAM research is conducted in groups of employees who are presented with a new piece of software and then use it for a period of time. After that TAM research is conducted (cf. e.g. [9], [15]). In our research — in order to get
Table 3: Cronbach Alpha reliability coefficients for two versions of the questionnaire

<table>
<thead>
<tr>
<th>Construct</th>
<th>Reliability 1</th>
<th>Reliability 2 after item removal (second group)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PU</td>
<td>0.63</td>
<td>0.73; 0.74; 0.66; 0.63</td>
</tr>
<tr>
<td>PEoU</td>
<td>0.64</td>
<td>0.80; 0.76; 0.82; 0.71</td>
</tr>
<tr>
<td>ATU</td>
<td>0.91</td>
<td>0.95; 0.94; 0.93; 0.91</td>
</tr>
<tr>
<td>BI</td>
<td>0.50</td>
<td>0.75; 0.65; 0.68</td>
</tr>
<tr>
<td>AU</td>
<td>0.36</td>
<td>0.63; 0.46; 0.68</td>
</tr>
<tr>
<td>scale</td>
<td>0.87</td>
<td>0.90</td>
</tr>
</tbody>
</table>

more general results — we asked of a system which is not connected to any organizational environment, and which was used by different participants in different frequency and circumstances (there were common characteristics though, since we have asked about CAPTCHA systems involved in e-mail accounts registration process). The other reason might be a bias in the very idea of this construct, i.e. self-reported usage — cf. [26, p. 6]. In our opinion the reliability of the AU construct in CAPTCHA research might be improved in two ways. One of them is to apply the procedure described above — i.e. to present participants with an instance of a CAPTCHA and invoke them to solve tasks for a while in defined environment and then perform TAM research. Another possible solution would be to measure AU concept in an experimental setting. As for the research presented in this paper we have decided to drop the construct from the further analysis. In our opinion this brings no harm to our scope of interests, since the most interesting part of the TAM model in CAPTCHA context would be PEoU, PU, ATU and BI. As a consequence, we will not consider 5th hypothesis stated in Section 2.1.

3.2. Model fit

Let us remind that the research hypotheses\(^1\) were, that positive relationships will be observed between:

- (H1) PU and BI;
- (H2) PEoU and ATU;
- (H3) PU and ATU;

\(^1\)Hypothesis 5 was dropped for the reasons pointed out in the previous subsection.
• (H4) ATU and BI;
• (H6) PEOU and PU.

Adequacy of TAM model for the collected data was performed only for the second group of data (gathered for the improved version of the questionnaire). The reason for this step were low reliability results for the first version of the questionnaire. Analysis results are shown in Figure 3 along with hypotheses numbers.

![Figure 3: Model fit results (\(\chi^2(9) = 144.8, p < .001; CFI = 1.00; NNFI = 1.03; RMSEA = 0, 90\% CI = 0; SRMR = .022\))](image)

Since strong PU and PEOU correlation was observed and confirmation analysis revealed stronger affect of one of them, we have decided to perform mediation analysis. We were interested if relation between PEOU and ATU is mediated by PU. Sobbel test was carried out and the results confirmed that PEOU influence on ATU is moderated with PU \((p < 0.001)\) — cf. Figure 4.

![Figure 4: Indirect effect Sobel’s test = .71 Z = 4.82 SE = .14](image)

Summary of the research results is presented in Table 4.
Table 4: Summary of research hypothesis findings

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>(H1) Positive relationships will be observed between PU and BI</td>
<td>not confirmed</td>
</tr>
<tr>
<td>(H2) Positive relationships will be observed between PEoU and ATU</td>
<td>confirmed, the effect is mediated by PU</td>
</tr>
<tr>
<td>(H3) Positive relationships will be observed between PU and ATU</td>
<td>confirmed</td>
</tr>
<tr>
<td>(H4) Positive relationships will be observed between ATU and BI</td>
<td>confirmed</td>
</tr>
<tr>
<td>(H6) Positive relationships will be observed between PEoU and PU</td>
<td>confirmed</td>
</tr>
</tbody>
</table>

3.3. Discussion and further research

The results obtained are consistent with previous TAM researches. The only exception is (H1), which was not confirmed — there were no direct relationship between PU and BI (such relationship was observed only for ATU and BI). The reason for that might be the research scenario (as described in the AU concept case). If the BI is measured long after the introduction of a system under research there might be a bias in results — cf. [15, p. 7]. In our opinion also CAPTCHA properties might be important here. The research concerning a class of systems (although with consistent characteristics), not a single software token.

One of the most interesting findings is that PU influence on ATU is stronger that PEoU. What is more, PEoU influence on ATU is mediated by PU. This seems to suggest that to build positive ATU among users, developers should concentrate not only on making CAPTCHA easier or more human-readable, but also on factors constituting PU concept. Let us remind that PU is defined as: “the degree to which person believes that using a particular system would enhance his or her job performance” [9, p. 320] In our research context this might be seen as the degree to which person believes that solving CAPTCHA would enhance his or her performance with e-mail account (or more broadly with a whole registration process and using this account). Four items used for this construct were the following:

- Using CAPTCHAAs increases security of web-pages that I use.
- Using CAPTCHAAs significantly increases the time of registering an e-mail account.
• When CAPTCHA task is difficult I tend to resign from registering an e-mail account.

• CAPTCHAs are responsible for decreasing the number of SPAM messages I receive.

The content of these items (especially the first and fourth one) suggests that one of the key issues about CAPTCHA is to make a user aware of the role and significance of this kind of task in the process of registration, and further using email account. This fact seems to be neglected by scientists investigating CAPTCHAs, as well as by the CAPTCHAs’ developers and e-mail account providers. When we check three very popular e-mail accounts providers — Gmail, Yahoo! and Hotmail, it appears that only on Hotmail, a user might find an information about what is this CAPTCHA picture for. But even there it is hidden under the ‘Help’ button. The explanation is following:

We ask you to enter the letters and numbers in a picture when we need to make sure that a person, not an automated program, is using Windows Live. The characters are drawn so that it’s possible for a person to recognize them, but very difficult for a program to. This helps us prevent automated programs from creating large numbers of accounts and sending spam. (https://signup.live.com/)

Neither Gmail, nor Yahoo! provides users with any kind of explanation. As for Gmail it is particularly awkward, since nearly all form input fields are explained or illustrated with an example, but there is none information to come with CAPTCHA picture.

Future research over this finding will concentrate on identifying necessary pieces of information that should be provided with CAPTCHA task to increase PU. Another interesting question for further research would be, if it makes the difference when a user know that solving CAPTCHA tasks results with some extra productivity, like for example in reCAPTCHA system case. The reCAPTCHA systems helps to digitize books, since its tasks consist of words unrecognizable by OCR machines. Human users recognize those words while solving reCAPTCHA. For detailed description see for example [3].

As it was mentioned in the text there are also further works needed to improve the questionnaire for measuring TAM constructs better in the field of CAPTCHA systems.


Appendix: Scales for measuring TAM constructs

<table>
<thead>
<tr>
<th>PU</th>
<th>Using CAPTCHAs increases security of web-pages that I use</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Using CAPTCHAs significantly increases the time of registering an e-mail account</td>
</tr>
<tr>
<td></td>
<td>When CAPTCHA task is difficult I tend to resign from registering an e-mail account</td>
</tr>
<tr>
<td></td>
<td>CAPTCHAs are responsible for decreasing the number of SPAM messages I receive</td>
</tr>
<tr>
<td>PEoU</td>
<td>Solving CAPTCHA is not comfortable</td>
</tr>
<tr>
<td></td>
<td>It is easy for me to understand how to solve a CAPTCHA task</td>
</tr>
<tr>
<td></td>
<td>It is easy for me to solve a CAPTCHA task correctly</td>
</tr>
<tr>
<td></td>
<td>I understand how to use additional options for a CAPTCHA task</td>
</tr>
<tr>
<td></td>
<td>Generally I see my interactions with CAPTCHA as positive (ble)</td>
</tr>
<tr>
<td>ATU</td>
<td>The idea of using CAPTCHA tasks in e-mail registration is bad/good</td>
</tr>
<tr>
<td></td>
<td>The idea of using CAPTCHA tasks in e-mail registration is foolish/wise</td>
</tr>
<tr>
<td></td>
<td>The idea of using CAPTCHA tasks in e-mail registration is useful/useless</td>
</tr>
<tr>
<td></td>
<td>The idea of using CAPTCHA tasks in e-mail registration is harmful/beneficial</td>
</tr>
<tr>
<td>BU</td>
<td>I will try to avoid e-mail accounts with CAPTCHA</td>
</tr>
<tr>
<td></td>
<td>If I would offer free e-mail accounts I would use CAPTCHA to secure them</td>
</tr>
<tr>
<td></td>
<td>I would like to solve CAPTCHA tasks if it would reduce SPAM messages I retrieve</td>
</tr>
<tr>
<td>AU</td>
<td>How often did you solve CAPTCHA tasks for the last six months?</td>
</tr>
<tr>
<td></td>
<td>How many tries do you need to solve a CAPTCHA task on a given web-page?</td>
</tr>
<tr>
<td></td>
<td>How often you fail to solve a CAPTCHA task on the first try?</td>
</tr>
</tbody>
</table>