



Research Journal of
**Information
Technology**

ISSN 1815-7432



Academic
Journals Inc.

www.academicjournals.com

CAPTCHA: A Survey of Usability Features

Rezvan Pakdel, Norafida Ithnin and Mohammad Hashemi

Department of Computer System and Communication, Universiti Teknologi Malaysia Skudai, 81310 Johor, Malaysia

Corresponding Author: Rezvan Pakdel, Department of Computer System and Communication, Universiti Teknologi Malaysia Skudai, 81310 Johor, Malaysia

ABSTRACT

The test that is generated and graded by computers in order to check the users to be human is Completely Automated Public Turing Test to Tell Computers and Humans Apart (CAPTCHA). CAPTCHA can protect online and free services from bots that can access to services in a huge numbers. Currently, there are several different types of CAPTCHA that have been proposed. However, most of proposed CAPTCHAs have Decaptchas that can read and recognize them. However, the main problem is not security of the current CAPTCHAs that is usability of it. The way of choosing features of a Captcha that previous researchers have used was not based on the fact of Human and Computer Interaction rules. This research will propose a sort of usability features based on what users choose and what needs to be considered of Human and Computer Interaction rules in order to develop a usable CAPTCHA.

Key words: Authentication system, CAPTCHA, human interaction proof (HIP), usability features

INTRODUCTION

CAPTCHA is a program has been used for preventing the robot to access the information and resources. A framework purposed to check the humanity in 1997. Later Von Ahn *et al.* (2004) developed a complete framework as Automated Turing Test. In 2000, A Human Interaction Proof (HIP) method was invented and called CAPTCHA by Von Ahn *et al.* (2004). CAPTCHA is stand for Completely Automated Public Turing test to tell Computers and Humans Apart. There are several web services and many other free services such as blogs, chat rooms, online polls, free e-mail accounts, search engines, password systems, etc., that are needed to be protected from robots not to be taken and wasted because of the limitation of the current resources.

Since 2000, there are many types of CAPTCHA developed to increase the security of CAPTCHA, however by raising the reliability of a CAPTCHA, its usability become decreased So, it has been needed to study existing reliability and usability elements of previous CAPTCHAs and find out that how they can affect reliability and usability of a CAPTCHA.

In fact there are many problems to recognize some CAPTCHAs and many of them seem to be annoying and silly but in fact in terms of using CAPTCHA it is understood that it has good ability to protect websites and their resources from robots. As there are many servers such as Gmail, Yahoo and Hotmail and so on, that provide free email services, there is a need to protect them from bots, which can create thousand accounts in a minute (Von Ahn *et al.*, 2004). CAPTCHA is a solution that can recognize human and bots. Also CAPTCHA has been used to prevent bots that create spam mail account in a huge numbers (Von Ahn *et al.*, 2004).

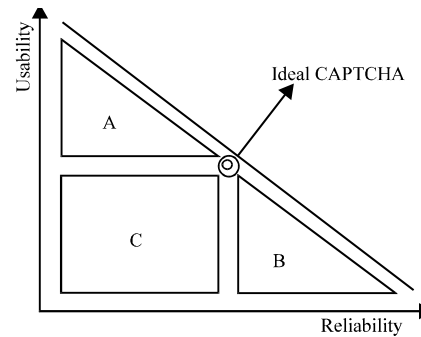


Fig. 1: Ideal usability and reliability of CAPTCHAs

There are many types of CAPTCHA that have been increasingly created to solve previous problems in terms of reliability or usability. In this study, we believe that the reliability and usability of a CAPTCHA should be balanced. The ideal reliability and usability of a CAPTCHA would be present in the Fig. 1.

In order to have an ideal CAPTCHA, the usability features of a CAPTCHA should be considered as much as the reliability features are.

The following parts are some literature review on applications of CAPTCHAs, different kinds of current CAPTCHAs and a survey of usability features.

BACKGROUND AND RELATED WORK

Application of CAPTCHA

Search engine bots: CAPTCHA can prevent search engine bots from reading web pages by using html tag. So bots cannot access to it (Chanamolu, 2009).

Protecting email addresses from scrapers: CAPTCHAs with an efficient mechanism can hide email address from Scrapers. So before showing your email address, User should solve this CAPTCHA. This is perfect solution that scrapers cannot see your email address (Chanamolu, 2009).

Protecting website registration: There are some companies such as Yahoo, Microsoft, Gmail, etc that provide free email services; each user can have one or more free emails. The issue here is attackers that by using some software programs can register thousand times for an email account every minute. CAPTCHA would be useful here to protect websites and free email services from attackers and bots.

Worms and spam: CAPTCHAs have efficient solution that can protect websites among bots and spam. It allows creating an email only by human not other software and then it accepts the new email.

Online polls: Online Polls have been grown rapidly since Internet was become popular everywhere. Moreover, it needs to be reliable sometimes. In 1999, a website started an online poll to choose the best computer science school; they knew that any user can vote more than one. So, they prevent users not to vote more than one by having users IP. Then two schools, CMU and MIT, found a way to use a program that can vote thousands at a time; a new contest was begun between the bots of these two schools (Chanamolu, 2009).

Table 1: Types of CAPTCHA

Type of CAPTCHA	Title
Text-based CAPTCHA	Early CAPTCHAs
	Improve CAPTCHAs
	Modern CAPTCHAs
	Animated CAPTCHAs
	ASCII CAPTCHAs
	Reverse CAPTCHAs
	Problem solving CAPTCHAs
Image-based CAPTCHA	Image recognition CAPTCHAs
	3D CAPTCHAs
Audio-based CAPTCHA	Audio
Video-based CAPTCHA	NuCAPTCHA

Preventing comment spam in blogs: Weblog designer are using some programs to add a huge number of bogus comments in the weblogs. The purpose of this operation is to get the higher ranks from search engine websites. The first solution to prevent these types of comment called comment spam was to register as a user to make a comment. However, these programs also can register themselves as a user and make comments. CAPTCHA is the best way to prevent comment spam in blogs.

Preventing dictionary attacks: Some types of CAPTCHA use a dictionary word inside the CAPTCHA. So some dictionary attackers can find the word inside the CAPTCHA by testing all words in the specific dictionaries. So the solution here is making a limitation for entering word (Pinkas and Sander, 2002). For example, each user can enter 4 or 5 times to solve a CAPTCHA, if user cannot be successful, another CAPTCHA will be appearing.






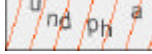


Types of CAPTCHA: There are four main categories of CAPTCHAs each of which has so many different schemes developed. Table 1 shows the type and category of current CAPTCHAs.

Text-based CAPTCHAs: The first CAPTCHA was Text based CAPTCHA. The text-based CAPTCHAs has been designed based on meaningful words or just series of characters which choose randomly from a specific database and by adding some noises or some mechanisms make it difficult to recognize by computers. When the user wants to access something that we need to be sure that the user would be human, CAPTCHA is shown and the user would fill the empty box with the word in CAPTCHA. If user fill box correctly, user can continue. Text based CAPTCHAs are shown in Table 2, are the most widely used CAPTCHAs (Zhu *et al.*, 2010).

Image-based CAPTCHAs: In the image-based CAPTCHA, the idea was to use images to make the CAPTCHA more difficult to recognize by bots. In the text mode, recognition is easily by some methods. But in image-based CAPTCHA because of having colours in all pixels and also having huge variety of meaningful images rather than texts and words, we will have better CAPTCHA (Zhu *et al.*, 2010).

In (Chew and Tygar, 2004), researchers created image labelling based CAPTCHAs by using labelled photographs. In this process, computer select six images with the same label and then user are asked about this label.

Table 2: Text-based CAPTCHA

Title	CAPTCHA	Title	CAPTCHA
Simple		Dashed	
Inclined		Inclined and dashed	
Mixed		Mixed and dashed	
Different colours		Different colours	

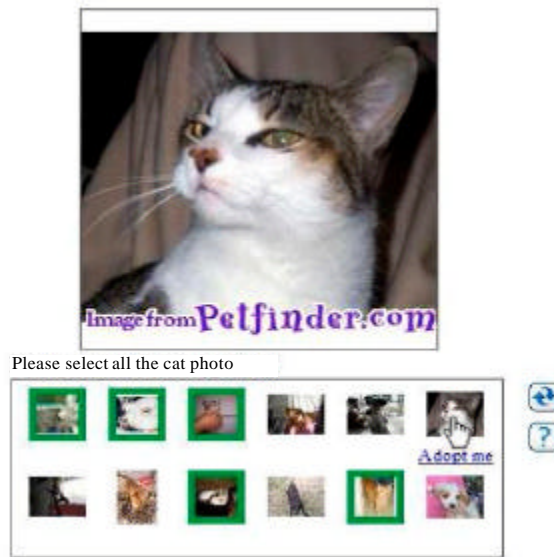


Fig. 2: Image-based CAPTCHA

In (Elson *et al.*, 2007), Asirra formally was introduced that is stands for animal species image recognition for restricting access (Asirra), a cat or dog labelling based CAPTCHA design. In this CAPTCHA asks users to choose a specific image. For example it asks the user to choose cat from 12 random chosen Asirra's GUI is illustrated in Fig. 2.

Audio-based CAPTCHAs: Audio-based CAPTCHA was designed firstly because of disability of some humans such as disability of vision and the problems related to eyes and so on. The human who wishes to access the protected resource must identify the text that is displayed correctly. So users with poor vision with some related disabilities cannot solve CAPTCHAs and alternative called the Audio CAPTCHA has been introduced. There are two type of model in this type of CAPTCHA. The first one that shows in Fig. 3, the users listen to a meaningful recorded voice with background noise. The second sound related with image (Zhu *et al.*, 2010). Also some existing audio CAPTCHA is highly time consuming (Bigham and Cavender, 2009). Current Audio-based CAPTCHAs have been broken by high-quality voice recognition and noise removal programs (Gupta *et al.*, 2009).

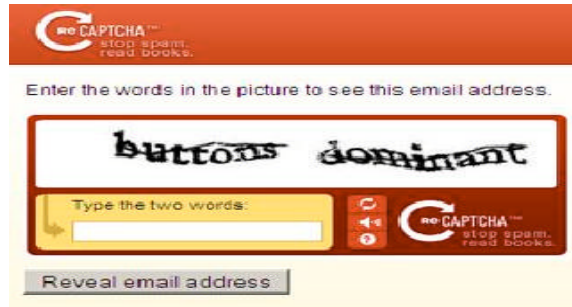


Fig. 3: Audio-based CAPTCHA

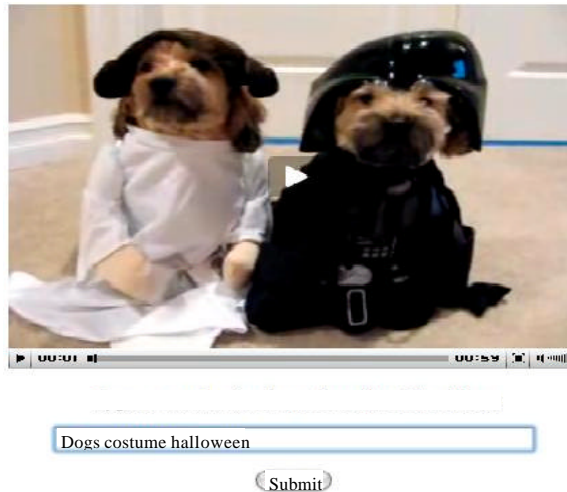


Fig. 4: Video-based CAPTCHA

Video-based CAPTCHAs: A video CAPTCHA using labelled video clips from YouTube is proposed (Kluever and Zanibbi, 2009). In this type of CAPTCHAs a user must type some words that have best describe in the video, Fig. 4.

USABILITY FEATURES OF EXISTING CAPTCHA SCHEMES

Table 3 shows a comparative study of all discussed CAPTCHA schemes in this study has been done based on our literature review and their proposed papers.

In terms of comparative study as a method to make a result of comparing the existing CAPTCHA schemes, Table 3 can be made. In Table 3, there are three main types of CAPTCHA that are text based CAPTCHA, Image Based CAPTCHA and Audio Based CAPTCHA. For Text Based CAPTCHA, six CAPTCHA scheme will be considered in order to compare to others. For Image Based CAPTCHA, there are two most important and well-known CAPTCHA scheme that are 3D and image recognition CAPTCHA.

The usability features can be categorized into three main categories that are complexity, content of CAPTCHA and presentation each of which has some sub category.

Table 3: Comparative study of existing CAPTCHA schemes

Type of CAPTCHA	Reliability		Usability		Complexity		Content of the CAPTCHA		Presentation		Range of age						
	Problem	Usability	Confusing characters	Language Independent	Non-meaning words	Long words	Guessable words	Obscene word	Difficult words	Font Type	Image size	Different colour	Web capability	5-12	12-18	18-60	Above 60
Text-based CAPTCHA																	
Early improve CAPTCHAs	X	✓	X	X	✓	✓	✓	X	X	✓	✓	✓	X	X	X	X	X
Modern CAPTCHAs	X	✓	X	X	X	✓	✓	✓	✓	✓	✓	✓	X	✓	X	X	✓
Animated CAPTCHAs	X	X	✓	X	✓	X	✓	X	✓	✓	✓	✓	✓	✓	X	X	✓
ASCII CAPTCHAs	X	✓	X	X	X	✓	✓	✓	X	X	X	X	✓	✓	X	X	X
Reverse CAPTCHAs	X	✓	✓	X	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	X	✓
Problem solving	✓	X	X	✓	X	X	X	X	X	X	X	X	✓	✓	X	X	✓
Image-based CAPTCHA	X	✓	X	✓	X	X	X	X	X	✓	✓	✓	✓	✓	X	X	✓
3D	✓	X	X	✓	X	X	X	X	X	✓	✓	✓	✓	✓	X	X	✓
Audio-based CAPTCHA	X	X	X	✓	X	X	X	X	X	✓	✓	✓	✓	✓	X	X	✓
Noise in background causes the distort sounds in audio CAPTCHAs Digits and letters read in a language are often not understandable to people who do not speak the language There is no standard graphical symbol for representing an audio CAPTCHA on a web page																	

✓: Yes X: No -:Not Mentioned

The range of age is a new features considered in order to have an ideal CAPTCHA that can be used by different age people. As it can be understood from the Table 3, most of the existing CAPTCHAs cannot satisfy the users of the age 5 to 12 years old and also above 60 years old. So it needs to be considered that users who are using the Internet services are in the different ages.

The content of CAPTCHA is also another category of usability problems. In terms of content problem, there are five different terms that need to be considered in order to design a new CAPTCHA that are non-meaning words, guessable words, difficult words, long words and obscene words. Most of current CAPTCHAs have not cared about these features and could not satisfy users to have all these features.

Confusing design can make a CAPTCHA usable but not reliable. There is a concept to have middle. If the design could be balanced in terms of reliability and usability, it could be said that this CAPTCHA is both reliable and usable.

RESEARCH METHODOLOGY

A questionnaire includes five parts and 43 questions have been prepared and piloted between 30 users in variety range of age, knowledge of CAPTCHA and education.

The purpose of piloting this questionnaire was to ask users about an ideal CAPTCHA that they would like to solve in order to access some resources in the Internet. It can be called user affinity of choices in order to design a usable CAPTCHA based on the usable features, which can be come from users themselves, NOT developers of CAPTCHAs.

Scaled question have been chosen for all questions to be easy to decide and answer by users at any knowledge and ages. However, in the first part of questionnaire, there is an introduction and a definition of what CAPTCHA is and for what purpose they are used.

Part two is about some personal questions includes educational information, age and their gender.

Part three includes some general questions about CAPTCHA like the users knowledge and their ability to solve a CAPTCHA.

In part four, there are 11 questions about the usability of current CAPTCHAs as well as their ideal CAPTCHA.

The most important part of this questionnaire is part five, which is some questions about designing a CAPTCHA and chooses some usable features that users like to have in the CAPTCHA they want to solve.

The result of the questionaire and the chosen features by users are explained in the following parts.

SURVEY RESULT

The questionnaires have been piloted between a group of non-academic and academic people such as students, professors and the people who access Internet and know about CAPTCHAs. One of the elements has been consider to evaluate our CAPTCHA Scheme is age. The previous CAPTCHA Schemes are not designed for different aged people. This study has found out that one of the problem of the current CAPTCHA is not to consider the knowledge and the age of users which lead this study to define five level for age of users as is shown in Fig. 5.

CAPTCHA is the authentication method used by men and women. A regard to this fact, this study considers users to be in fair of number of male and female users as are shown in Fig. 6.

Our users are academic and non-academic people who have different educational knowledge and degree as are shown in Fig. 7.

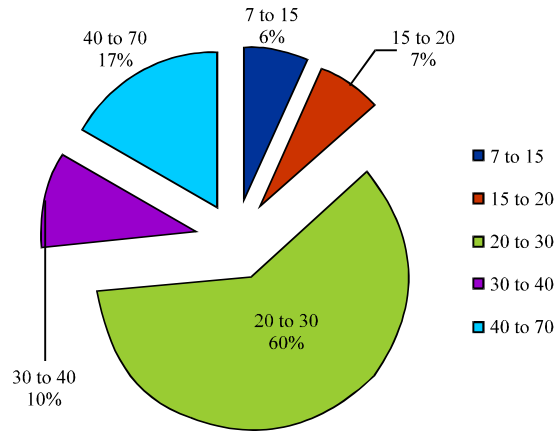


Fig. 5: Participant's range of age

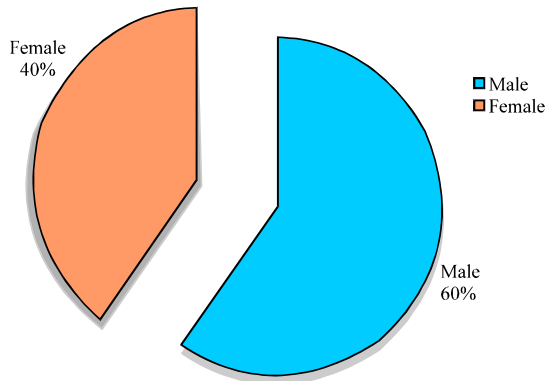


Fig. 6: Participant's gender

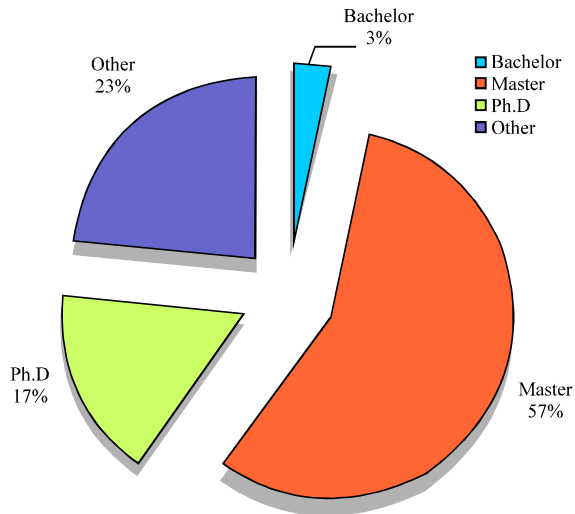


Fig. 7: Participant's educations

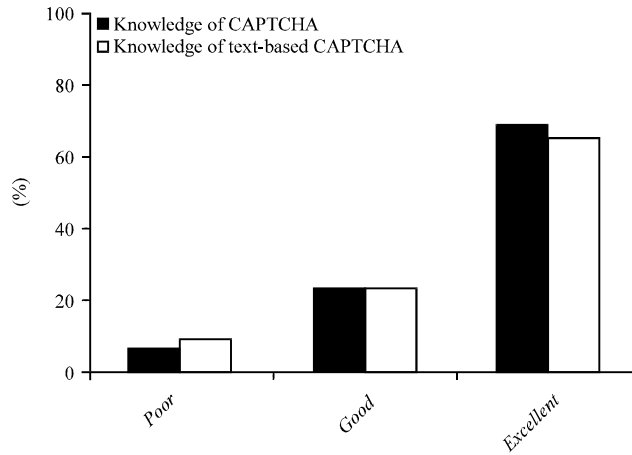


Fig. 8: Participant's knowledge of CAPTCHA

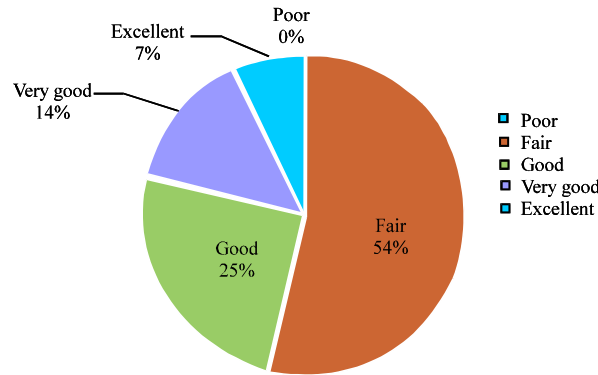


Fig. 9: Participant's ability of solving current CAPTCHAs

To evaluate user's feedback it is needed to control their knowledge and be sure to get the appropriate feedbacks. The knowledge of our participants is shown in Fig. 8.

Our participants claim their ability to solve a Text-based CAPTCHA and the other available CAPTCHA that are using currently. As it is shown in Fig. 9, 54% of our participants considers themselves a fair CAPTCHA solver. CAPTCHA solver is a person who needs to recognize or solve a CAPTCHA in order to access a specific resource. This 54% has probably difficulty at solving current CAPTCHA schemes and would help this study to make a usable CAPTCHA.

Also our survey participants claim that they are not fast at solving a CAPTCHA in the current version of it. As it is obvious in Fig. 10, 83% of them considers their speed of solving current CAPTCHAs fair and poor. The most amazing thing is that there is no one claims to be very good or excellent at solving a CAPTCHA fastly or even normally.

Our survey participants have been asked about the possibility of solving current CAPTCHA scheme have been faced for the people in different age. The result has shown that the people in the age of 5 to 12 years old have difficulty to solve a CAPTCHA. Also they claim that the current CAPTCHA is hard to solve by old people because of the characteristics of CAPTCHA's scheme such as using text or the way of entering solution. The people aged 18 to 60 has no problem at solving a CAPTCHA at any scheme. Regard to this fact, the other group of people should be considered as the target of designing a new scheme of CAPTCHA. The survey also asked

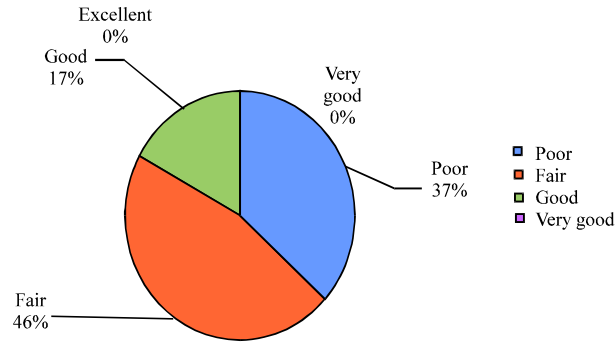


Fig. 10: Participant's speed of solving current CAPTCHAs

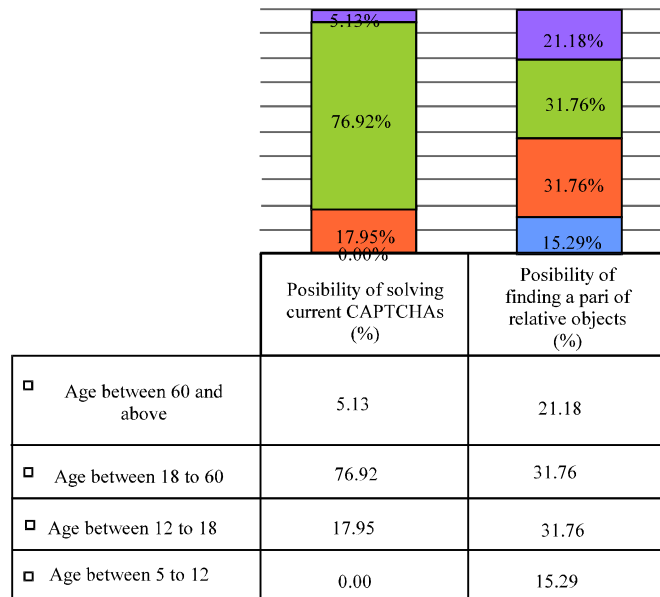


Fig. 11: Current CAPTCHAs vs. Pair of relative objects

participants to rate the possibility of finding a pair of relative objects that this study would consider for the final proposed CAPTCHA scheme. Finding a pair of object like chair and table or pen and paper is suitable for people in different age and knowledge. There are four group of people consider to solve a CAPTCHA. The first group is the people aged 5 to 12 years old; second one aged 12 to 18; the third group has 18 to 60 years old; the last group is the people who are older than 60 years old.

Each group should have 25% satisfactory to show that they have no problem at finding a pair of objects. Fortunately all these groups have a percentage near to 25%. It is clear the most sensitive group is the people with the age of 5 to 12 years old because of their experiences and knowledge. Figure 11 shows the possibility of solving current CAPTCHAs and finding a pair of relative objects for the people in four different scales of age.

As it is mentioned before that there are four types of CAPTCHA includes text-based, audio-based, video-based and Image-based CAPTCHA. Our participants have been asked the suitability of these types of CAPTCHA for four groups of people with four different scales of age. Figure 12 show that Text-based CAPTCHA is not suitable for the people aged 5 to 12 years old and

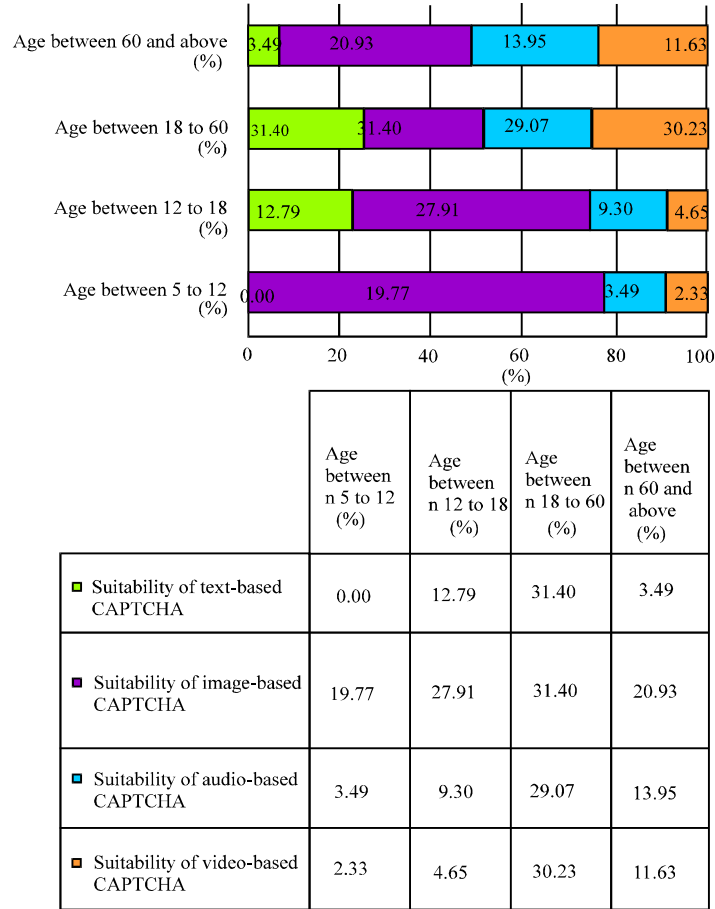


Fig. 12: Suitability of current CHAPTCHAs

above 60 years old. Also Video-based and Audio-based CAPTCHA are not suitable for the people aged 5 to 12 years old. The only group that could be consider as the most normalize one with the normal percentage is Image-based CAPTCHA. Based on this part of our survey, Image-based CAPTCHA has been chosen as one of the features of a usable CAPTCHA.

Figure 13 shows the satisfactory of some elements used in the current CAPTCHA schemes extracted from related works of this study.

As it is shown in Fig. 13, there are some elements used to design current CAPTCHAs that this study finds them the characteristics that made a CAPTCHA unusable. Our participants' idea was not to use them in a usable CAPTCHA. They have decided not to use noises, bad color and complexity in a usable CAPTCHA. Also they argue that the reason of unsatisfactory of text-based password was to see offensive words, confusing characters and long length texts. In the Image-based CAPTCHA they don't like to have a rotation and an image in bad quality condition.

Another part of our questionnaire was about their opinion about the current CAPTCHAs using in the Internet and some software in order to prevent robots to access the resources already dedicated for human. In our point systems which considered a good point for satisfactory and a bad point for unsatisfactory, Image-based CAPTCHA got the best rank and point as it is shown in Fig. 14. Most of our participants have chosen unsatisfactory for the other schemes and they have chosen Text-based CAPTCHA as the worst one.

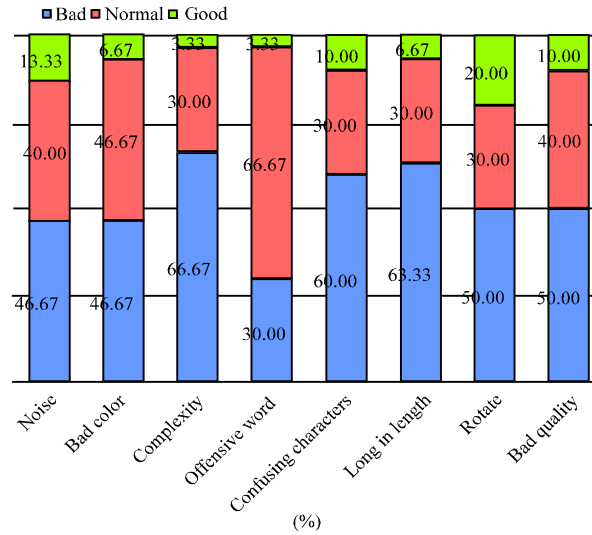


Fig. 13: Satisfactory of current CAPTCHA's features

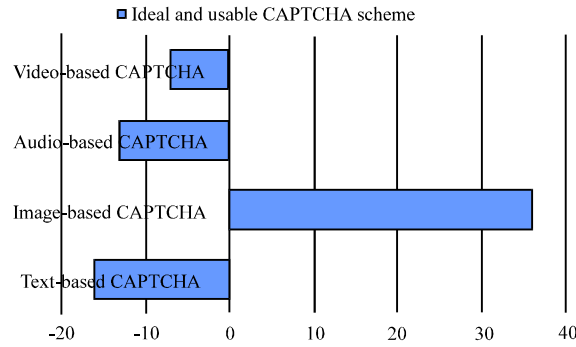


Fig. 14: Satisfactory of current CAPTCHA schemes

In some cases, some Image-based CAPTCHA has been designed as a hybrid scheme, which has text and image together. Figure 15 shows that the user does not like to have alphanumeric characters inside an Image-based CAPTCHA and they prefer Images rather than Texts. The other feature should be considered in a new usable CAPTCHA is to use only images.

The other difficulty that the other studies have shown in order to satisfactory of a CAPTCHA was the way of entering the solution of a CAPTCHA. In the current CAPTCHA schemes users need to enter the solution in a text box by entering characters of a keyboard. Using a keyboard is the thing that people everyday have been faced, however during working in Internet users tend to use a mouse more than a keyboard. Moreover, by thanks to new invention of touchable devices, touch and click are so popular in the high rate of usage. Figure 16 shows that our participants tend to use mouse rather than keyboard in order to enter the solution of a usable CAPTCHA. Therefore the other feature for a new usable CAPTCHA is to design how a user can enter the solution by using mouse.

Finally, as there are some CAPTCHA schemes that are multi option questions and are more likely than others, it would be possible to choose a solution between some options. In order to have

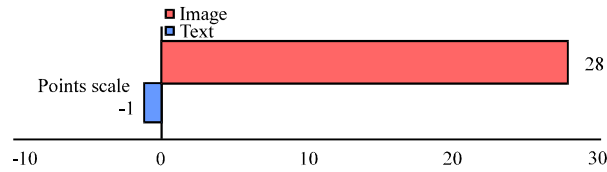


Fig. 15: Image vs. Text in a CAPTCHA

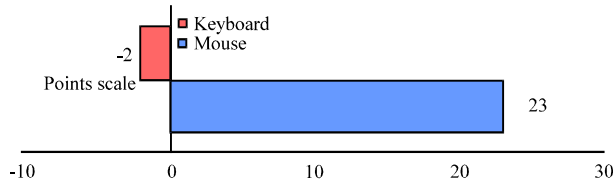


Fig. 16: Keyboard vs. Mouse

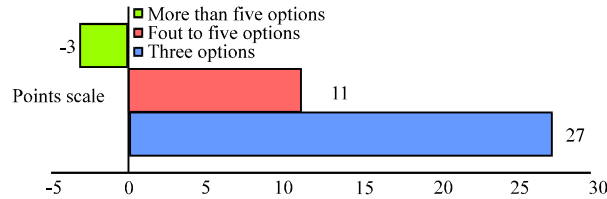


Fig. 17: Number of options in CAPTCHA

a user affinity of choice, our participants have been asked how many options they would have to choose the correct solution better. Figure 17 is showing that they have chosen to choose the solution of a CAPTCHA between three options.

CONCLUSION FOR PROPOSED FEATURES OF CAPTCHA

This study is based on user affinity of choices and care about what users like to have and see inside a CAPTCHA; then a CAPTCHA can be considered a usable CAPTCHA scheme. This study has concerned about the usability features of current CAPTCHA schemes and extracts the features of the previous CAPTCHA schemes in order to evaluate them. After evaluation, some new features were added and users were asked about them. The result of this survey is the features that users chose as usable features.

All results were without considering the age, knowledge, race, sex and the educational information of people. All selected features should be considered completely in order to make a powerful and usable CAPTCHA scheme.

Image-based CAPTCHA is the most usable scheme users have chosen as the best scheme. The selected scheme should not use text and alphanumeric characters. Designers of a usable CAPTCHA should not use noise, bad color, images with bad quality, complex ideas and questions and rotation. Also they should design a usable CAPTCHA that might be solved by using only a mouse meaning that entering the solution of a usable CAPTCHA should be done by using a mouse. In that scheme, designers would consider not more than three options to choose. It means that the question should not have more than three options.

ACKNOWLEDGMENT

I would like to thank my mother and father who have been supporting me in whole my life and all I have is theirs.

REFERENCES

- Bigham, J.P. and A.C. Cavender, 2009. Evaluating existing audio CAPTCHAs and an interface optimized for non-visual use. Proceedings of the 27th International Conference on Human Factors in Computing Systems, April 4-9, 2009, Boston, MA., USA., pp: 1829-1838.
- Chanamolu, Y., 2009. An orientation based image CAPTCHA. Master Thesis, Ohio State University, Nigeria.
- Chew, M. and J. Tygar, 2004. Image recognition CAPTCHAs. Proceedings of the 7th International Information Security Conference, September 27-29, 2004, Palo Alto, CA., USA., pp: 268-279.
- Elson, J., J. Douceur, J. Howell and J. Saul, 2007. Asirra: A CAPTCHA that exploits interest-aligned manual image categorization. Proceedings of the 14th ACM Conference on Computer and Communications Security, October 28-31, 2007, Whistler, Canada, pp: 366-374.
- Gupta, A., A. Jain, A. Raj and A. Jain, 2009. Sequenced tagged captcha: Generation and its analysis. Proceedings of the IEEE International Advance Computing Conference, March 6-7, 2009, Patiala, Punjab, India, pp: 1286-1291.
- Kluever, K.A. and R. Zanibbi, 2009. Balancing usability and security in a video CAPTCHA. Proceedings of the 5th Symposium on Usable Privacy and Security, July 15-17, 2009, Mountain View, CA., USA., pp: 1-11.
- Pinkas, B. and T. Sander, 2002. Securing passwords against dictionary attacks. Proceedings of the 9th ACM Conference on Computer and Communications Security, November 18-22, 2002, ACM Press, Washington, DC., USA., pp: 161-170.
- Von Ahn, L., M. Blum and J. Langford, 2004. Telling humans and computers apart automatically. *Commun. ACM*, 47: 56-60.
- Zhu, B.B., J. Yan, Q. Li, C. Yang and J. Liu *et al.*, 2010. Attacks and design of image recognition CAPTCHAs. Proceedings of the 17th ACM Conference on Computer and Communications Security, October 4-8, 2010, Chicago, IL., USA., pp: 187-200.