International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 05 | May 2021 www.irjet.net p-ISSN: 2395-0072

Captcha for Visually Impaired People: A Review

Sakshi Baheti[1], Hardik Shah[1], Rijul Sherathia[1], Prakash Waghamode[2]

[1]Student, [2]Professor, School of Computer Engineering and Technology, MIT World Peace University Kothrud, Pune 411038

Abstract - *Nowadays the Internet users are from different ages and groups. The disabled people also use the Internet. Some websites are especially created for disabled people. Many Internet sites offer services for human users, but unfortunately some computer programs are designed which abuse these services. As a result, some systems named CAPTCHA (Completely Automated Public Turing test to tell Computers and Human Apart) [3] have been introduced to tell apart human users and computer software. In this paper, a new CAPTCHA method is introduced which blind people can use. In this method, a simple shape-based problem is created according to predefined patterns and converted to speech using a Text-To-Speech (TTS) system. Then the sound is played for the user and he/she must draw the answer of the question. Because answering this problem requires recognizing the speech, understanding the problem, and drawing the problem, only a human user can answer this question and present computer programs are unable to solve it. In addition, answering the question is easy for blind people, because the question consists of a number of natural language sentences and the answer is a number, which can be entered easily.*

***Key Words*: Web application, captcha, machine learning, visually impaired, motion capture, natural language processing, text-to-speech**

1.INTRODUCTION

Today lots of daily activities such as education, shopping and mailing are done through the Internet. By rapid growth of the Internet and the easy access to it, a great deal of private and personal information is available on the web.

Various methods are presented to overcome the above problem. The goal of these methods is distinguishing human users from computer programs. These methods are done automatically by computer programs since examination of a large amount of registration forms by human forces needs a great deal of time and expense. CAPTCHA (Completely Automatic Public Turing Test to Tell Computer and Human Apart) are systems which are used to tell humans and machines apart automatically. These systems are based on Artificial Intelligence (Al) topics [3]. They are similar to the Turing test, but they differ in that the judge is a computer. The goal of these systems is to ask questions which human users can easily answer, while current computer programs cannot. CAPTCHA systems also have other applications such as prevention from sending spam. Considering today's scenario, nowadays the Internet is not only for special groups of people, but also people from any age and different groups are using the Internet. There are many websites for children on the Internet and the children using the Internet for various activities such as entertaining, educating, etc. The elderly people communicate and chat with their children and relatives using the Internet. In addition to regular people, the disabled peoples are also using the Internet. Deploying only visual captchas creates a considerable obstacle to a certain group of users, such as the visually impaired, colour-blind, near sighted users. This necessitates the need for more accessible captchas.

Current CAPTCHAs rely on superior human perception, leading to CAPTCHAs that are predominantly visual and, therefore, unsolvable by people with vision impairments. Audio CAPTCHAs that rely instead on human audio perception were introduced as a non-visual alternative but are much more difficult for web users to solve. But the conventional Audio CAPTCHA alone cannot solve this problem, thus we need a better version of the audio captcha.

We need to keep in mind a few things while building a captcha for the blind: [3]

- Must have audio output
- Must be multilingual
- Questions must be easy
- Response ways should be effective

Audio CAPTCHAs have been shown previously to be difficult for blind web users. Sauer et al. found that six blind participants had a success rate of only 46% in solving the audio version of the popular reCAPTCHA, and Bigham et al. observed that none of the fifteen blind high school students in an introductory programming class were able to solve the audio CAPTCHA guarding a web service required for the course.

Thus, with these observations in mind, we have thought of a modified and effective way of presenting the audio captcha.

This will be multilingual for ease of understanding, simple questions for clarity, ML and motion-based captcha for security and supported by the latest backend technologies for faster response.

2. LITERATURE REVIEW

Paper [2] states that CAPTCHA methods can be generally divided into three groups: OCR-Based, Visual Non-OCR-Based and Non-Visual. Optical Character Recognition (OCR) programs are used for automatically reading the texts, but

they have difficulty reading texts printed with a low quality and can only recognize high-quality typed texts that use common standard formats. However, the defects of the OCR systems can be taken as an advantage by changing the picture of a word so that it can only be recognized by a human and not by any OCR system

2.1 OCR - Based Captcha

In OCR-Based methods [2], the image of a word with distortion and various pictorial effects is shown to the user and he/she is asked to type that word. Due to the presence of various pictorial effects, the computer will encounter problems in the recognition of the word and only a human user can recognize the word. Examples of these methods are: Persian/Arabic Baffle Text and Gimpy. But these methods usually result in dissatisfaction of users. On the other hand, efforts such as Non-OCR have been made for breaking these methods.

Most CAPTCHAs on the web today exhibit the following pattern: the solver is presented text that has been obfuscated in some way and is asked to type the original text into an answer box. The technique for obfuscation is chosen such that it is difficult for automated agents to recover the original text but humans should be able to do so easily. Visually this most often means that graphic text is displayed with distorted characters.

Text-based Captcha [3] is the most widely used captcha in web application. It is an image of distorted text/numbers and addition with some background noise or clutter. The content is generated randomly either text or alphanumeric. The user asked to identify the distorted letters or numbers whatever displayed in the captcha challenge and entered them. It requires a large question bank. It is uncomplicated to solve for visual users, but it becomes very difficult to read for blind users.

Visual CAPTCHAs, in paper [1], are perceived as a whole and can be viewed even when focus is on the answer box. Once focusing the answer box, solvers can continue to look at visual CAPTCHAs, edit the answer that they provided, and verify their answer. They can repeat this process until satisfied without pressing any keys other than those that form their answer. Errors primarily arise from CAPTCHAs that are obfuscated too much or from careless solvers.

Image-based CAPTCHAs [7] are designed by using various image objects. The user has to recognize a specific image to pass the test. Sometimes the images are provided with tags and the user is asked to identify the correct image and enter appropriate words in the box given or asked to click on a specific image to prove as a human user. The advantage of image-based CAPTCHA is that pattern recognition is a very hard AI problem and thus it becomes difficult to break this test using pattern recognition technique.

2.2 Non-OCR Based Captcha

In [4], there are Non-OCR Based methods which are more comfortable for users than OCR-based methods. These methods mainly based on the features of multimedia such as pictures and sounds and usually using methods like small puzzle games. Examples of these methods are PIX and Collage CAPTCHA. One of the main categories of non-visual methods are sound based CAPTCHA methods. In these methods, a word is said and the user must type the word. They are based on weaknesses of speech recognition systems. These systems are usually used beside other CAPTCHA methods, especially OCR-Based methods, for disabled people.

However, some of the Internet websites are especially designed for disabled persons. Also, these websites need protection against computer programs which try to use website resources. But common CAPTCHA systems are usually more difficult for disabled people than some of the Non-OCR Based methods which can be used by disabled people.

According to [7], generally a puzzle-based CAPTCHA can either be a graphics-based puzzle or a mathematical puzzle. In picture-based puzzles, the picture is divided into some pieces and provides these pieces randomly. Each piece of the picture will have a piece number. The user needs to arrange these pieces properly by following piece numbers to form a complete original picture, the puzzle based on mathematics is hundred percent effective and it can be incorporated into the login process and online form registration in the websites for ensuring legitimate access. The user needs to solve that math puzzle provided in order to get legal permission to access web contents.

Audio-based CAPTCHAs [8] are usually used as a complement for text-based CAPTCHAs. Many popular websites such as eBay, yahoo and Microsoft use both visual and audio CAPTCHAs. An audio CAPTCHA generally picks a random sequence of letters or numbers; renders them into a sound clip; includes some level of distortion; and then presents the recording to the users. The user is asked to type the contents of the recording. In one type of audio CAPTCHAs, known as spoken CAPTCHA, the users are required to repeat the test instead of typing it. This feature makes this CAPTCHA also suitable for blind users.

Paper [8] also describes motion-based CAPTCHAs in which a movie or animation is presented to the users and they are asked to recognize an action, animated word or image in the movie. This CAPTCHA is convenient for users. In addition, since the required processing time in this CAPTCHA is relatively high, it is more secure. However, the high loading time can be a disadvantage from a usability viewpoint. Another disadvantage is requiring a large database of animations. Finally, the term “hybrid CAPTCHA” has been selected for a CAPTCHA that is a combination of different types or designed for special purposes.

2.3 Visual Non-OCR Based & Non-Visual Captcha

Visual Non-OCR-Based and Non-Visual methods which are easier to pass for users than OCR-based ones[2]. Visual Non-OCR-Based methods use the drawbacks of computer vision systems such as their difficulties in identifying the type of an object in an image. For example, in Collage CAPTCHA, the image of some objects with distortion is shown to the user and he/she is asked to click on a certain object. One of the main categories of non-visual methods are sound based CAPTCHA methods. In these methods, a word is said and the user must type the word. They are based on weaknesses of speech recognition systems. These systems are usually used beside other CAPTCHA methods, especially OCR-Based methods, for disabled people. Disabled people usually have problems with CAPTCHA methods, because most CAPTCHA methods are designed for and tested by non-disabled persons.

Paper [4] studied and examined Drawing Captcha and Collage captcha (Non-OCR Based captcha) methods which can be used by disabled people.

Drawing Captcha method is for devices like PDA (Personal Digital Assistant) which uses a stylus. In this method, numerous dots are drawn on a screen with noisy background and the user is asked to connect certain dots to each other.

In the Collage Captcha method, the images of some different objects are chosen. Then some effects such as rotation are done on the images and they merged to create a single image. This image is shown to the user and he/she is asked to click on a certain object (for example on the image of the apple). Collage CAPTCHA requires a database of labeled images. Creating this database is expensive and requires a lot of time. It is an easy CAPTCHA method for users, because in this method the user must find the object image whose name is shown. In addition, this method may have a high rate of random passing. If the images of different objects can be easily separated, and the number of different objects is N, the probability of passing the test with a random answer can be 1/N. Disabled people such as hearing or sight impaired persons can use this method because it uses images without any distortion. In addition, mobility impaired persons can use this method easily because it requires only one click.

Paper [4] suggests an Non-OCR-based CAPTCHA method that is designed for blind people. In this suggested method, a simple mathematical problem is created according to predefined patterns and converted to speech using a Text-To-Speech (TTS) system. Then the sound is played for the user and he/she must answer the question. A computer requires the following abilities to answer the question:

- Recognition of the question using Speech Recognition systems.
- Understanding the meaning of the question.
- Solving the problem and answering the question.

Since it is difficult for computer programs to succeed in doing any of these operations, only a human user can answer the question.

Paper [2] proposes a new structure for CAPTCHA systems. In this method, a word is selected and then converted into an audio file using a Text-To-Speech (TTS). Then the audio is played for the user and he/she is asked to repeat the word which he/she hears. The user response is then analyzed by two modules. The first module identifies synthesized voices and hence identifies computer users. The second module is a speech recognizer and checks the user response to be the desired word. This module prevents recorded human voice attacks. The main contribution of our system is that it uses drawbacks of computers in both speech recognition and speech synthesis, while previous sound-based methods use only speech recognition drawbacks. So, its tests are easier for human users and more difficult for computers, in comparison to previous sound-based CAPTCHA.

3. RESEARCH GAPS

On reviewing [1], [3], [4] papers we figured out a few important things, which are not present very widely, and we would be planning to highlight and develop those features when we implement the project.

In [3] paper, we observed various captcha methods that were used for generation of captcha on websites and forms. The type which grabbed our attention was the Audio captcha as it is an effective method to be presented before visually impaired people. After reading through it, we understood that the conventional audio captcha was missing out on the capabilities of ML-AI, which is at its peak in today's time and is ever growing. It is extremely simple for a ML algorithm to capture the audio in the audio Captcha shown on the site and convert it into text and feed into the input. Thus, even if this method is suitable for the visually impaired people we cannot rely on the old audio captcha methods for differentiating between human and robot.

In [4] paper, we noticed a really effective and secure way of using Captcha that was Drawing CAPTCHA. This method is for devices like PDA (Personal Digital Assistant) which uses a stylus. In this method, numerous dots are drawn on a screen with noisy background and the user is asked to connect certain dots to each other. In view of the problems that computers face in recognizing the dots from the noise, only a human user can. With all the noise added and the approach where the user needs to connect dots to crack the Captcha would make it easier for the system to identify if a bot is trying to crack it. After our analysis, we decided to modify this way of CAPTCHA so that it is user friendly for visually impaired people as well along with maintaining the level of security it provides. With combination of ML Image recognition and motion capture in JavaScript, we can

probably figure out a functionality to modify the above approach for drawing captcha.

Paper [1] revolved more around research that identified the groups of people that were targeted for the captcha. This included visually impaired and people with normal eyesight. Various surveys were to understand how do the target audience react to different methods of captcha. Participants were first presented with a questionnaire asking about their experience with web browsing, experience with CAPTCHAs and the level of difficulty frustration they present, as well as demographic information. They were then asked to solve 10 visual CAPTCHAs and 10 audio CAPTCHAs (for sighted participants) or 10 audio CAPTCHAs (for blind participants). Each participant was asked to solve one problem randomly drawn from each CAPTCHA type, and the CAPTCHA types were presented in random order to help avoid ordering effects.

From the results of the above survey, we were able to understand that blind people show much frustration towards visual Captcha compared to audio captcha and that is obvious. However, this tells us that audio method is a suitable method for the blind and thus we need to maintain the same level of ease in audio CAPTCHA but modify it such that it is safer and more effective. Also, we saw that there is an opportunity to explore more features for showing the multi-linguality of the captcha audios.

Along with this we also noticed that there were no proper instructions provided to the users (visually impaired) on solving the captcha. We propose to build our solution where the target audience will get a clear and simple explanation on the quick steps they need to follow to crack the captcha and that will be available in their mother tongue.

All the previous work done on this topic were definitely turning points in the advancement of CAPTCHA but with the growth of ML-AI, we need to dig in more and find ways such that these methods retain themselves on today's web sites but are more secure and in a way more effective towards visually impaired along with general public.

4. PROPOSED ARCHITECTURE

4.1. OBJECTIVES

- To assist a visually impaired person to use and crack captcha on web applications.
- To combine the strengths of NLP (Natural Language Processing) to make multilingual captcha for better understanding.
- To classify the user as bot or human.
- To track the user keystrokes.
- To track the user cursor movements. Capturing the motion.
- To consume user input as speech if required.
- To build multiple choice captcha for ease of the user so that they can check the preferred captcha option for recognition.

4.2. MODULES

This application aims at simplification for visually impaired people while accessing captcha authentication. System will provide features of language selection in which the user is comfortable. A web application will be built using Flask at backend and HTML CSS features for UI along with the bootstrap library. This will be helpful for providing a strong backend and user-friendly UI for the users. Along with that, we will also be utilizing some features of Natural Language Processing like text to speech and speech to text conversion. We will also be using multiple corpuses for providing a multilingual experience to our target audience.

The proposed application uses machine learning algorithms for image classification and helps distinguish between a bot and human. One of the approaches for questioning will also be drawing captcha which will explicitly use motion capture and then feed that image to the backend ML model [5] which will classify and verify the image.

A simple use case diagram explains all the scenarios when a user interacts with the application in Fig: 1.

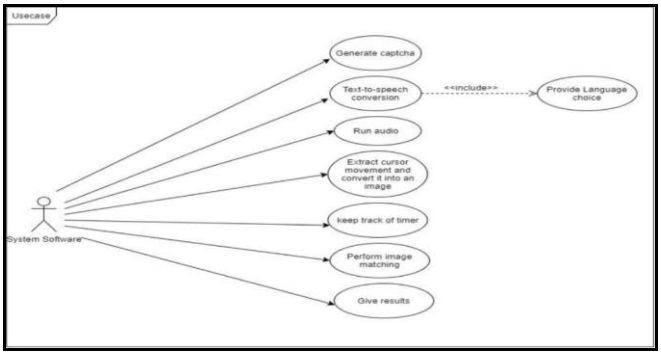

Fig: 1

Module 1: Captcha Generation

In the Captcha Generation module, the user will get different types of Captcha questions which they will have to answer and if answered correctly the ML model will classify them as human response and the app can then redirect to the required form or page. Given a question on the screen, the app will pick up different sets of questions from our database and present it on the screen. For a normal user, it will be easier to just read the question and fill in the answer while for the visually impaired, we have tried to maintain the same level of simplicity but have kept the way of communication as audio which is discussed in the next module.

Module 2: Audio output with multilingual options

As direct reading of the captcha question on the screen is not possible for visually impaired users and thus, we need audio to let the users understand what is the exact question.

The audio output can be generated using the text to speech conversion in NLP. [6] Along with this, NLP also provides various methods and using different corpuses, we can generate audio output in different languages. The main purpose of implementing this functionality is increasing the scope of our project. For various geographies, we can implement this method and people can hear the audio output in their mother tongue.

Module 3: Drawing captcha and keystroke detection

As we have audio output for blind users, we must also take care of the inputs that we will take and make it simple for the target audience to give input. Audio output will help users communicate which keys they need to press to perform a specific function or navigate around. Along with that, we also have one captcha questioning type where users will be asked to draw a simple shape. Thus, to capture the shape, we will use the motion capturing in JavaScript where the cursor movement will be captured. Once these are captured and converted to an image, it is sent to the backend ML model which will compare it with existing test data and classify if it was a human response or bot response. Thus, it will help us maintain the security of the system while keeping it simplistic.

A sequence diagram in Fig: 2 depicts a desired flow of user inputs and responses from the system.

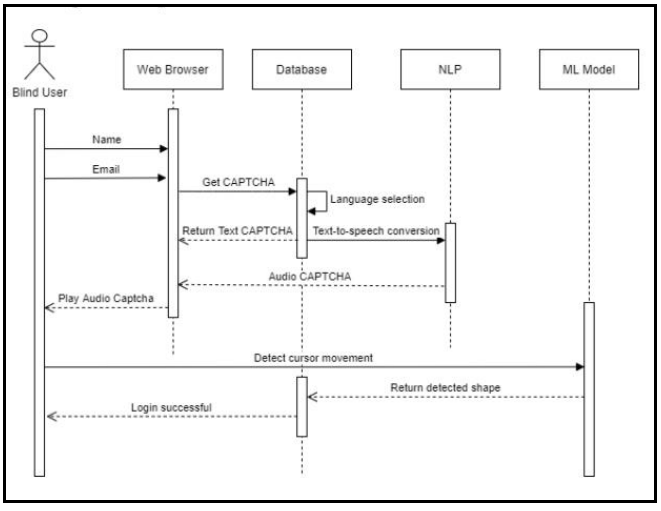

Fig: 2

5. REQUIREMENT

5.1. HARDWARE REQUIREMENTS

Monitor: CRT or LCD monitor
Keyboard: Normal or Multimedia
Mouse: Compatible mouse or Touchpad
Audio Output Device

5.2. SOFTWARE REQUIREMENTS

NLP: Natural Language Processing, or NLP for short, is broadly defined as the automatic manipulation of natural language, like speech and text, by software. The study of natural language processing has been around for more than 50 years and grew out of the field of linguistics with the rise of computers.

Machine Learning: Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves.

Python: An open-source general-purpose interpreted, high-level programming language can be used to create web apps, desktop applications, games, data science applications, and a variety of other items.

Flask: Flask is a lightweight WSGI web application framework. It is designed to make getting started quick and easy, with the ability to scale up to complex applications. It began as a simple wrapper around Werkzeug and Jinja and has become one of the most popular Python web application frameworks.

JavaScript: JavaScript, often abbreviated as JS, is a programming language that conforms to the ECMAScript specification. JavaScript is high-level, often just-in-time compiled, and multi-paradigm. It has curly-bracket syntax, dynamic typing, prototype-based object-orientation, and first-class functions.

6. CONCLUSIONS

In this paper, we have studied various types of CAPTCHA. A brief review has been carried out on the Audio Captcha and lists out the limitation in sound captcha in diverse methodologies. A huge scope for research exists in outlining new and novel CAPTCHA procedures that are easy to use, require less server handling and offer enhanced security control against bots and feasible for disabled as well as normal people.

In this paper, we propose a new structure for CAPTCHA systems. In this method, a task performing question is selected and then converted into an audio file using a Text-To-Speech (TTS). Then the audio is played for the user and he/she is asked to perform the task which he/she hears using the movement of the cursor. The ML model designed to detect the movement of the cursor will validate the performed task according to the question asked and allow

the visually impaired people to successfully login with any trouble.

7. REFERENCES

[1] Jeffrey P. Bigham and Anna C. Cavender. 2009. Evaluating existing audio CAPTCHAs and an interface optimized for non-visual use. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09). Association for Computing Machinery, New York, NY, USA, 1829–1838.

[2] S. Shirali-Shahreza, H. Abolhassani, H. Sameti and M. Hassan, "Spoken captcha: A captcha system for blind users," 2009 ISECS International Colloquium on Computing, Communication, Control, and Management, Sanya, 2009, pp. 221-224, doi: 10.1109/CCCM.2009.5270464.

[3] K. Aiswarya, K. S. Kuppusamy, "A Study of Audio Captcha and their Limitations", Department of Computer Science, Pondicherry University, Puducherry, India, April 2015

[4] M. Shirali-Shahreza and S. Shirali-Shahreza, "CAPTCHA for Blind People," 2007 IEEE International Symposium on Signal Processing and Information Technology, 2007, pp. 995-998, doi: 10.1109/ISSPIT.2007.4458048.

[5] S. Ray, "A Quick Review of Machine Learning Algorithms," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 2019, pp. 35-39, doi: 10.1109/COMITCon.2019.8862451.

[6] P. M. ee, S. Santra, S. Bhowmick, A. Paul, P. Chatterjee and A. Deyasi, "Development of GUI for Text-to-Speech Recognition using Natural Language Processing," 2018 2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech), Kolkata, 2018, pp. 1-4, doi: 10.1109/IEMENTECH.2018.8465238

[7] Prakash S. Bodkhe, Dr. P. N. Mulkalwar (2017) CAPTCHA Techniques: An Overview: 2nd National Conference Recent Innovations in Science and Engineering (NC-RISE 17) IJRITCC, 2017

[8] Roshanbin, Narges & Miller, James. (2013). A survey and analysis of current CAPTCHA approaches. Journal of Web Engineering. 12. 1-40.