Towards a Universally Usable CAPTCHA

Graig Sauer, Harry Hochheiser, Jinjuan Feng, and Jonathan Lazar
Department of Computer and Information Sciences, Towson University
8000 York Road, Towson, MD 21252
{gsauer1,hhochheiser,jfeng,jlazar}@towson.edu

ABSTRACT

CAPTCHAS are widely used on websites to differentiate between humans and computers. Audio CAPTCHAs provide an alternative interface that is accessible to blind users and users with low vision who are unable to see visual CAPTCHAs. We present a small user study suggesting that one existing audio CAPTCHA is highly error prone and time consuming. We propose a new CAPTCHA solution that combines audio sounds (e.g., bird chirping) and visual images (e.g., bird). Preliminary user tests of this new solution have been highly encouraging..

General Terms

Design, Experimentation, Security, Human Factors

Keywords

CAPTCHA, visual impairment, security, universal usability

1. INTRODUCTION

CAPTCHAs are widely used on websites for the purposes of differentiating between humans and computers. Typical CAPTCHAs are visually based interfaces that ask the user to identify characters in distorted text.. These visual CAPTCHA products are not very usable for visually impaired users. It has also been documented that humans with near perfect vision struggle to interpret the distorted text in these CAPTCHAs (false positives) [2]. The following are three examples of CAPTCHA products currently being used on the internet today:

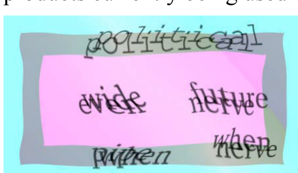


Figure 1, An example of the Gimpy CAPTCHA[7]

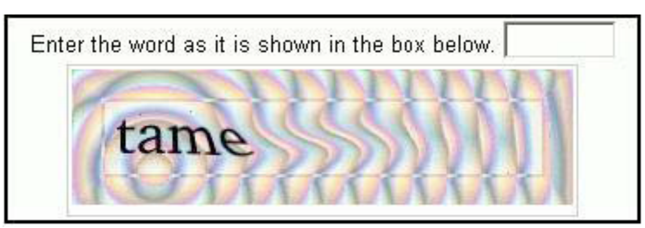


Figure 2, An example of the EZGIMPY CAPTCHA[7]

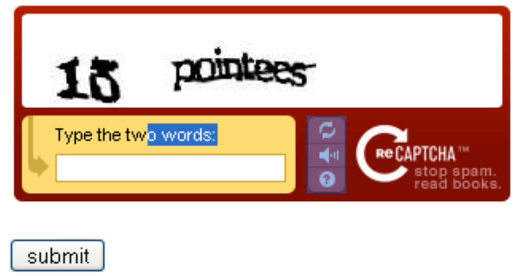


Figure 3, An example of the ReCaptcha CAPTCHA[9]

The first example is GIMPY, which is a twisted text CAPTCHA product. Users must identify the words in the image [7]. EZGIMPY is another distorted text CAPTCHA product in which users have to identify multiple words [7]. ReCaptcha asks users to identify the two words in the image. ReCaptcha also presents the user with an audio version, which asks users to type a string of 8 digits [9].

Table1. Widely recognized CAPTCHA products.

CAPTCHA Name	Type of CAPTCHA	Answer Type
Gimpy [7]	Character, Recognition	Text Box
EZ-Gimpy [7]	Character, Recognition	Text Box
BONGO [3]	Image, Anomaly, Recognition	Text Box
PIX [3]	Image, Anomaly, Recognition	Text Box
ASIRRA [5]	Image, Recognition	Text Box
IMAGINATION [6]	Image, Recognition	Dropdown List
audio/image [4]	Image, Recognition, Sound	Dropdown List

Table 1.provides an overview of existing CAPTCHA solutions. Those solutions can be grouped into five categories:

Character Based: This category means that a string of characters is presented to the user. This string can contain either words or random alphanumeric characters.

Image Based: Images or pictures are presented to the user. This is normally in the form of an identifiable real-world object, but can also be presented in the form of shapes (BONGO). The task is to identify the object shown in the picture.

Anomaly Based: Users are asked to determine which object, or character, or shape does not belong in a set of images displayed on the screen..

Recognition Based: The user needs to determine what is being presented to them. In the case of a character based and recognition based CAPTCHA the user needs to identify and input the character string that is presented to them.

Sound Based: The user is presented with an audio version of a CAPTCHA. The user listens to the audio file and inputs their answer. A sound based CAPTCHA can be presented in two formats, the first is the "spoken words or numbers" and the second would be sounds related to an image.

Developers have recognized the accessibility shortcomings of the visual CAPTCHA and have begun research into this area. One major shortcoming of CAPTCHAs based on spoken text or numbers is that the audio has to be distorted to defeat the use of automated speech recognition to solve the challenges. Because of this distortion it becomes difficult even for a human to differentiate between the distortion and the valid data [8].

We are unaware of any research that has been done on the audio CAPTCHA to measure its accessibility/usability for blind or low vision users. In this context we use the term accessibility to mean not having barriers that block people, and we use the term usability as being able to achieve the goal. Although such a system may be accessible, it is not clear that it is particularly usable. Our goal is to conduct a study that will not only provide insight on the accessibility and usability of audio CAPTCHAs for vision impaired users, but to also inform design of a more robust and accessible CAPTCHA system.

2. Accessibility/Usability Study

Our usability study was designed to address several accessibility and usability concerns with current audio CAPTCHAs.

- Can users comprehend the CAPTCHA audio with distortion in it?
- Can users easily remember the numbers contained in an audio CAPTCHA challenge to use to answer the CAPTCHA?
- Can users use their screen reader to find the controls that start and operate the audio CAPTCHA? Is there any interference between the screen reader (JAWS) output and the audio CAPTCHA? Does this interference prevent users from successfully completing the CAPTCHA?
- Are there any frustrations that arise from any of the above factors?

2,l Study Design

We implemented a webpage based on the ReCAPTCHA product developed by Carnegie Mellon University. This CAPTCHA product offers both a visual distorted text CAPTCHA and an audio CAPTCHA. For the audio CAPTCHA the user is presented with an audio clip in which eight numbers are spoken by various individuals. Background noise is also inserted into this audio clip to make it harder for hacker Bots to break the CAPTCHA. The user is then supposed to fill in a form with those eight numbers and hit a submit button. Upon hitting the submit button they are presented with either a "Correct" or "Incorrect" reply.

The computer that the participants were using was pre-configured with the JAWS screen reader, with an audio output rate that was selected based on our prior experience working with JAWS users. Participants were allowed to use external aids of their choosing, but were not allowed to adjust the audio output rate.

Each participant was asked to solve the audio CAPTCHA six times, including one practice trial. Tasks were evaluated on both time and correctness of completion. The practice attempt was designed to give the users some comfort with the site, eliminating the possibility of disorientation. At the end of each participant's session they were asked to fill out a short demographics questionnaire about their experience during this study.

2.2 Demographics

Six individuals (3 males, 3 females), ranging in age from 28 to 54 participated in the study. One participant (user #5) had partially impaired vision, the rest were blind. Participants had a range of experience with computer use and with the JAWS screen reader – Table 2.

Table2. Participants computer experience.

Participant	Years of Computer Use	Hours of Computer Use Per Day	Jaws Experience on a scale of 1-10 (1 Being the lowest)
1	18	4	7
2	4	6-8	7
3	15	7-8	8
4	10	7	9
5	20	10	1
6	20	8	10

2.3Results

Figure 4 shows the number of correctly completed attempts for each individual. Figure 5 shows the average completion times for each user's correct attempts. Figure 6 shows the average completion times for each users failed attempts

Table3. Techniques used by participants to solve the CAPTCHA.

Participant	Technique

1	Used a Braille Note Taker to document numbers as they were spoken
2	Memorized numbers as they were spoken then type them in at the end
3	Memorized numbers as they were spoken then type them in at the end
4	Used a Word document to document the numbers as they were spoken
5	Participant with residual vision, typed numbers into the submit box as they were spoken
6	Memorized numbers as they were spoken then type them in at the end

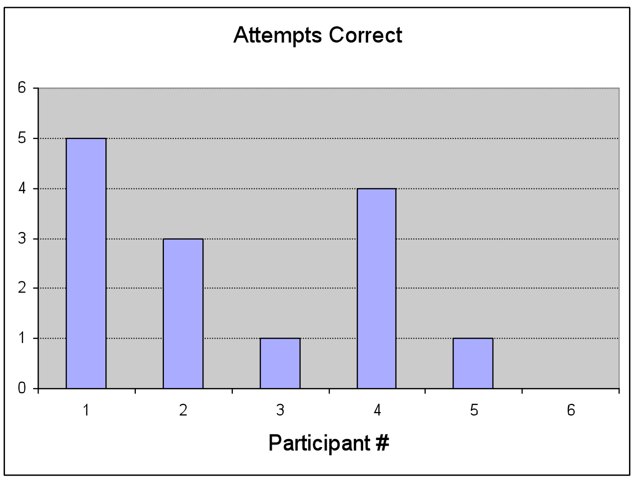


Figure 4, Graph of participants and the number of attempts they got correct.

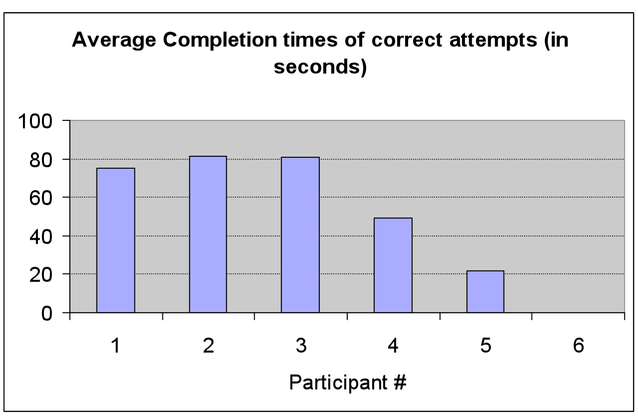


Figure 5, Graph of participants and the average amount of time taken on each correct attempt.

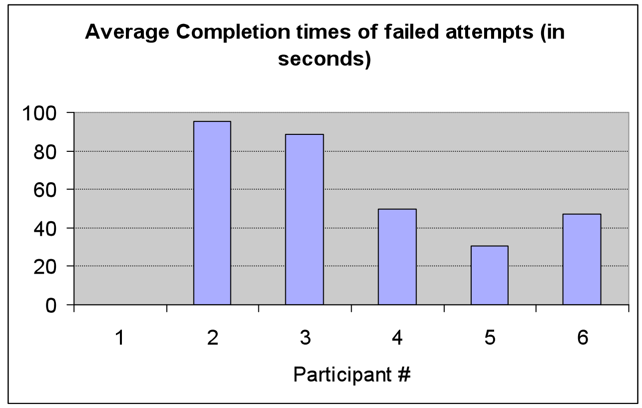


Figure 6, Graph of participants and the average amount of time taken on each failed attempt

The average amount of time spent overall on correct attempts was 65.64 seconds. The average time overall spent on failed attempts was 59.56 seconds.

These results show that there was a wide range of correct answers. Some people were able to complete all of the tasks correctly, but others were not able to get any of their attempts correct. There was also a very wide range of average times taken to complete this CAPTCHA. Some were able to complete it in an average time as low as 30 seconds were others took almost as long as 90 seconds to complete the task. Participants that used some sort of external aid (Braille note -taker or MS Word) were much more successful then those that did not use these aids.

2.4 Discussion

The study participants who were most successful relied upon external aids, perhaps indicating that the audio CAPTCHA passes the threshold of accessibility but does not achieve the goal of usability. One possible explanation for this result is that the audio CAPTCHA imposes more cognitive load than an average human user can handle, a situation called cognitive overload. Consequently, users have to utilize external tools to release the cognitive load of the task.

The audio CAPTCHA fell well short of the 90% success rate [2] that human users are supposed to have for solving a CAPTCHA. This study showed that a visually impaired user could only solve the CAPTCHA at a rate of 46%, 44% less then what is supposed to be.

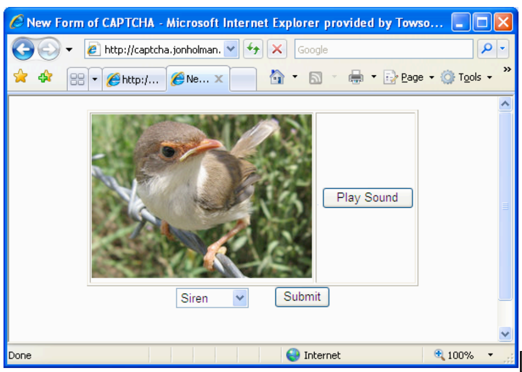
The average amount of time taken to correctly solve an audio CAPTCHA of 65.64 seconds is greater then the 51 seconds that is suggested as the time to complete a CAPTCHA[1]. The average time was 14.64 seconds greater then that time. Although we do not expect blind user times to rival sighted user times we do think that the average times should be a bit closer.

Five of the six participants complained about the clarity of the audio CAPTCHA. Two of these five admitted to guessing on numbers that they did not understand.

3. Towards an Accessible CAPTCHA

Ideally, CAPTCHA design should be informed by the values of universal usability: tools should support a broad range of users of differing backgrounds and abilities. Current CAPTCHA systems create a separation between their visual and audio CAPTCHAs. . The audio CAPTCHA is essentially a distinct system with a completely independent development and maintenance path. Alternatively, the visual and audio CAPTCHA can be joined products into one single system in which the audio is directly related to the visual elements that are presented to the user. We feel that this type of CAPTCHA will be more accessible for users with visual impairments, as well as having possible benefits of easy adaptation for different languages and cultures.

The image/audio CAPTCHA [4] combines pictures of familiar objects with sounds associated with these objects (Figure 6). If the user is visually impaired he is able to solve the CAPTCHA based on the sound, and if the user is hearing impaired he is able to solve the CAPTCHA based on the picture. Preliminary user testing was positive.



[4]

Figure 6, Image/audio CAPTCHA concept

Current challenges involve scaling the size of the image/sound database to a level suitable for production use. Construction of a large sample of easily distinguishable sound is challenging. We are also working on extending the tool to require identification of multiple sounds/images, instead of the one used in the initial prototype. Finally, the use of single, well-defined audio files creates the possibility using of checksums or signatures to defeat the CAPTCHA. We plan to solve this problem in our next prototype.

Our expanded prototype increases the number of image/sound combinations to 30 and allowed for multiple sound/image combinations. The user will have to get all correct in order to pass the CAPTCHA. Runtime concatenation of sounds with intervening silence was added to help discourage checksum/signature based attacks.

Conclusions and Future Research

Planned further development of the prototype includes adding non-audible noise to the sound files on the fly in order to deter potential checksum/signature attacks. User tests will evaluate the usability and accessibility of the prototype.

4. ACKNOWLEDGMENTS

Our thanks to The National Federation Of the Blind for assisting us with recruiting participants.

5. REFERENCES

[1] Travis Schluessler, Stephen Goglin, and Erik Johnson. Is a Bot at Control? Detecting Input Data Attacks. *Proceedings of the 6th ACM SIGCOMM workshop on Network and system support for games*, 2007.

[2] Kumar Chellapilla, Kevin Larson, Patrice Simard, and Mary Czerwinski. Designing Human Friendly Human Interaction Proofs (HIPS). *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2005.

[3] Louis Von Ahn, Manual Blum and John Langford. Telling Humans and Computers Apart Automatically. *Comm. Of the ACM*, 47(2):57-60, 2004.

[4] Jonathan Holman, Jonathan Lazar, Jinjuan Feng, and John D'Arcy. Developing Usable CAPTCHAs for Blind Users. *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, 2007.

[5] Jeremy Elson, John Douceur, and Jared Saul. Asirra: A CAPTCHA that exploits Interest-Aligned Manual Image Categorization. *Proceedings of the 14th ACM conference on Computer and communications security,* 2007.

[6[Ritendra Datta, Jia Li, and James Wang. IMAGINATION: A Robust Image-based CAPTCHA Generation System. *Proceedings of the 13th annual ACM international conference on Multimedia*, 2005

[7] Greg Mori and Jitendra Malik. Recognizing Objects in Adversarial Clutter: Breaking a Visual CAPTCHA. In *Computer Vission and Pattern Recognition*, 2003.

[8] Andy Schlaikjer, A Dual-Use Speech CAPTCHA: Aiding Visually Impaired Web Users While Providing Transcriptions and Audio Streams. CMU-LTI-07-014. http://www.lti.cs.cmu.edu/

[9] ReCAPTCHA:Stop Spam Read Books (2007), Retrieved on May 7, 2008 at http://recaptcha.net/