Audio CAPTCHA: Usable Accessibility + Usable Security

Dr. Charudatta Jadhav
Tata Consultancy Services Ltd
Intersil Building, SEEPZ, Andheri
(East), Mumbai- 400096
Charudatta.jadhav@tcs.com

Sumeet Agrawal
Tata Consultancy Services Ltd
Intersil Building, SEEPZ, Andheri
(East), Mumbai- 400096
Sumeet.a@tcs.com

Abstract
Visual CAPTCHAs are designed to be unreadable by machine, hence, are inaccessible to assistive technology users. Traditional Audio CAPTCHAs are theoretically accessible to visually impaired (VI) users but are often hard to solve with low success rate, i.e., below 50% and can be exploited by automated attacks. Either of the attributes - security or usability gets compromised in order to make an accessible CAPTCHA. In this paper, we propose a novel concept for audio CAPTCHA, which aims to meet all the three attributes - accessibility, security, and usability. We evaluated the usability, security and accessibility aspect with the visually impaired and mainstream users. The task success rate was 100% for mainstream users and 83% for the screen reader users.

Author Keywords
Accessibility; Accessible and Secure CAPTCHA; Usable Security; Usable Accessibility.

ACM Classification Keywords
Human-centered computing: Accessibility-Accessibility design and evaluation methods.

Security and privacy: Security services

Introduction

CAPTCHAs (Completely Automated Public Turing test to tell Computers and Humans Apart) are commonly used in websites to determine whether or not the user is human [1]. As visual CAPTCHAs (based on image) are not accessible to people using assistive technology like screen readers, the audio CAPTCHA functionality was created. An audio file having voice pronunciation of random digits/alphabets with background sound effects is presented to the user to identify and type. Prior research literature highlights, humans find traditional audio CAPTCHA hard and obfuscated to solve as they sometimes struggle to recognize the letters e.g. letters which tend to sound similar such as like T and D, B and P with the background noises present [2]. Also from the security aspect, these audios can be decoded by high-quality ASR and variety of advanced machine learning algorithms [3] with high success rate [2]. Thus, proving that accessibility, security, and usability of these kinds of audio CAPTCHA doesn't go hand-in-hand, as either of them gets compromised. This trade-off between the attributes gave us the motivation to come up with a new approach of presenting audio CAPTCHA.

Our larger goal is to design an Inclusive CAPTCHA which is accessible to every user with all kind of abilities without compromising security and usability. In this paper, we discuss the concept behind the aural form of the Inclusive CAPTCHA.

The proposed aural form of Inclusive CAPTCHA

The audio CAPTCHA is built upon the concept of presenting the audio file having a real-life scenario along with the contextual question. Figure 1 is the working example diagram of the aural form of Inclusive CAPTCHA.

The CAPTCHA generation module consists of libraries of audio files of real-life scenario based on the range of themes for example conversation in school, traffic, railway stations, zoo, and parks. For each theme, the library consists of sub-libraries of audio files pertaining to main dialogues, environmental noise, and background noise and background speech. The real-time audio mixer component randomly picks two or more audio files from each sub-library of the corresponding selected theme and merges them in an overlapping manner to generate a single audio media file. On receiving the request from the user, the audio file and corresponding contextual question to solve the CAPTCHA is presented to the users. On submitting the response to the question, the CAPTCHA self-learning solution determination module is an expert system which uses information retrieval technology strategies to match the response input by the user and stored response of the question taking care of issues like spelling errors, incomplete responses, contextual metonyms, synonyms, and variants. As for the example shown in Figure 1, the acceptable user response for the question asked can be 'math/mathematics/sum/plus/add/addition' which only a human can interpret and answer.

The question asked in the CAPTCHA is purely based on the context of the school theme, making it easier for the user to relate and comprehend as they are prone to experience the scenario in their day to day life presented in the audio file, thus taking care of the usability aspect. The level of possible combinations for each theme, the possible variants and the overlap of the audio files, enables addressing the security aspect.

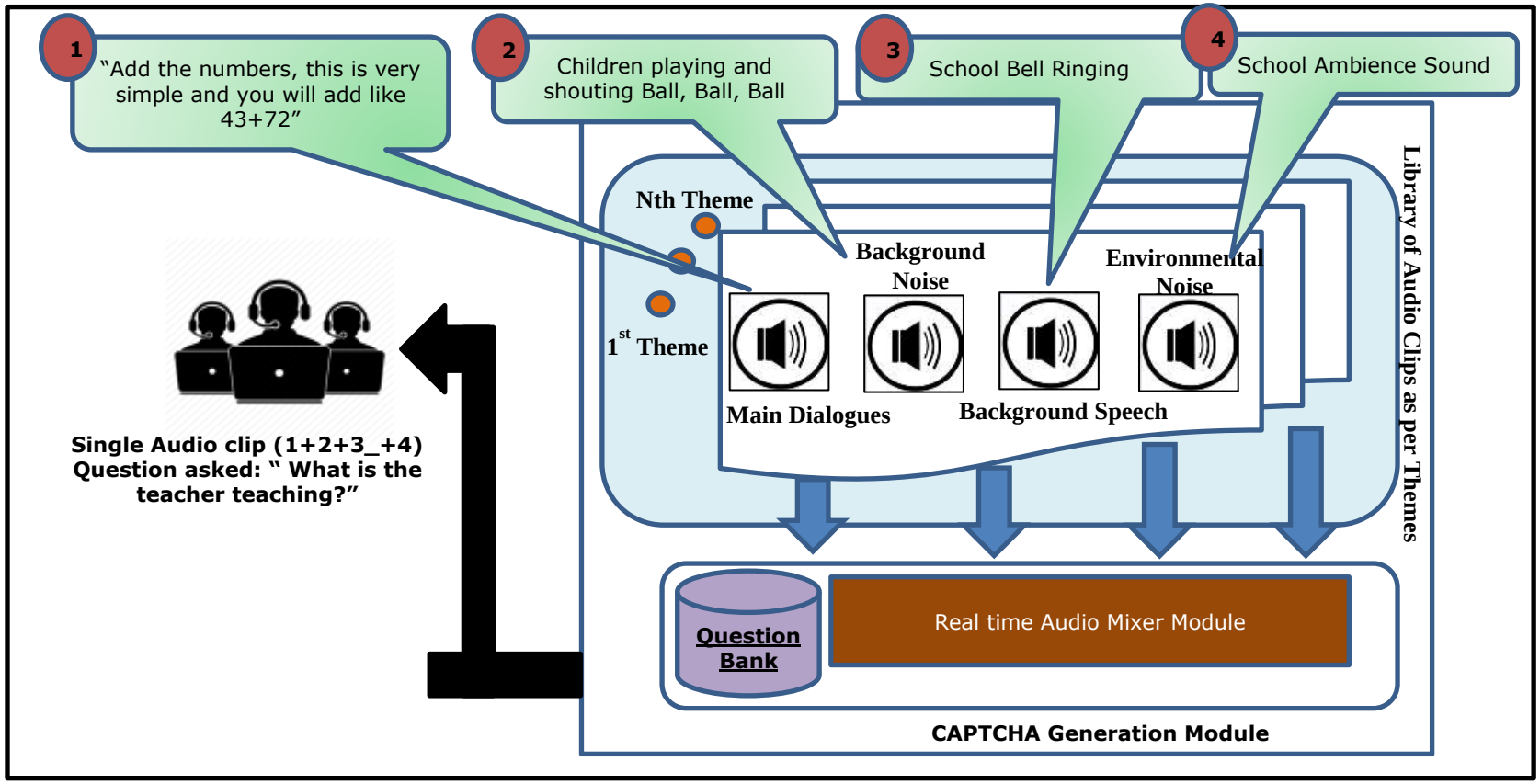

Figure 1: System Diagram and working example of audio form of Inclusive CAPTCHA

Evaluation Study:

The evaluation study was conducted in 2 parts. The first part of the evaluation was done to find out whether the audio form of Inclusive CAPTCHA can be broken by an attacker, and the second part of the study was done to evaluate the usability aspect of the CAPTCHA. A total of 119 participants took part in this user study where 24 were screen-reader users (partially blind and completely blind), and 95 were non-disabled (mainstream) users who were recruited by crowdsourcing.

Security Analysis: There are primarily two ways to break the security of the Audio CAPTCHA.

- Involving a person to capture all of the audio files with relevant answers database and an automated script to simulate the answers.
- Decode the audio CAPTCHA with high-quality ASR and information extraction algorithms

As per the first approach if a human or a bot was created to solve the CAPTCHA, listening the all possible audio files recording the answers into a database, preparing combinations and simulating it, will take a lot of time and is practically difficult.

For the automated algorithmic solving first the program will have to convert the speech into text to get the audio data and then later input it to the classifier trained on audio data to predict which topic or the answers of the question which may arise within the themes of the audio. To verify the same, we used IBM Watson the best speech recognition AI platform and cognitive computing system. **Error! Reference source not found.Error! Reference source not found.**, below shows sample of the transcripts obtained for the constructed audio CAPTCHA files by ASR. The results indicate that the transcripts derived from ASR were not anywhere close to the actual scripts of the audio and it would be difficult for the classifier even to predict the theme used in the audio CAPTCHA file.

Table 1: Comparison of ASR engine and Actual audio files transcripts

Theme Name	***Transcripts obtained from ASR***	***Actual Scripts of the Audio***
Temple Environment	He couldn't refrain greenness	Where are you going, Kiran? I am going to the Temple, John.

User Performance Analysis: Figure 22 below shows the analysis of the successful and failed attempts analysis. The term 'successful attempts' represent the number of users who have passed the CAPTCHA test in the first attempt and 'failed attempts' represents the number of users who have been unable to solve the CAPTCHA in the first attempt. Nevertheless, they were able to complete the CAPTCHA task after two or more attempts. From the graph below 88% of the mainstream (non-disabled) users were able to pass the CAPTCHA challenge in the first attempt. Approx. 84% (20 out of 24) of the screen-reader users were able to pass the CAPTCHA challenge in the first attempt which we believe is a remarkable improvement on the 46% pass rate given in [5] and 70% pass rate for reCAPTCHA given in [3].

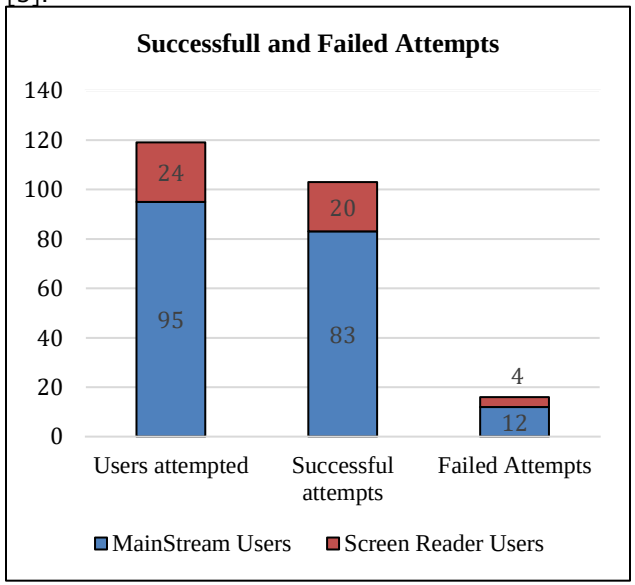

Figure 2: Graph of successful and failed attempts of the users solving the aural for of the CAPTCHA challenge

Conclusion:

The analysis and the result of the study showed the following:

- Security test results show aural form of the Inclusive CAPTCHA is hard to crack and decode.
- Accessibility testing reported no issues.
- Usability test shows positive responses with task success rate of 100%, and 83% of the screen reader users participated were successful in the 1st attempt without any initial training.
- Screen Readers users were able to complete the task with an average time of 38 seconds with a minimum of 14 seconds which is less as compared to previous studies **Error! Reference source not found.**.

With the larger goal of creating Inclusive CAPTCHA, we will be extending the concept used in the creating audio CAPTCHA to the visual form of the CAPTCHA to ensure it is accessible, usable and secure at the same time for all the range of people with disabilities and mainstream users. Then we plan to extend the evaluation study with the user group like hearing impaired, cognitive impaired, motor impaired users.

We hope to present our additional analysis and experiment results during the presentation at EMPOWER 2018.

References

[1] https://en.wikipedia.org/wiki/CAPTCHA, accessed 4 Apr 2017

[2] Aiswarya K. & KuppusamyK. S. et al 2015. "Study of Audio Captcha and their Limitations", International Journal of Science and Research, 4(4), 918-923. DOI= https://www.ijsr.net/archive/v4i4/SUB153197.pdf

[3] Tam, J., Simsa, J., Hyde, S., & Ahn, L. V. et al.2008. "Breaking audio captchas". In Advances in Neural Information Processing Systems (pp. 1625-1632). DOI= https://papers.nips.cc/paper/3443-breaking-audio-captchas.pdf

[4] Sauer, G., Holman, J., Lazar, J., et al. 2010. "Accessible Privacy and Security: A Universally Usable Human-Interaction Proof tool." Universal Access in the Information Society, 9, 3. 239-248. DOI=https://link.springer.com/article/10.1007/s10209-009-0171-2

[5] Sauer, G., Lazar, J., Hochheiser, H., et al. 2010."Towards A Universally Usable CAPTCHA" Human Interaction Proof: Evaluation of Task Completion Strategies. ACM Transactions on Accessible Computing, 2, 4. 1-32. DOI= https://cups.cs.cmu.edu/soups/2008/SOAPS/sauer.pdf