# Distinguishing Humans from Automated Programs by a novel Audio-based CAPTCHA

Seyhmus Yilmaz
Department of Computer Engineering
Duzce University
Turkey

Sultan Zavrak
Department of Computer Engineering
Duzce University
Turkey

Huseyin Bodur
Department of Computer Engineering
Duzce University
Turkey

## ABSTRACT
CAPTCHA (Completely Automated Public Turing Test to Tell Computer and Humans Apart) has become a ubiquitous guard utilized to prevent exploitation in web services like account registration. They are universally secure measure to distinguish real users from automated programs by using computer-generated tests that should be easy for users to solve but should be hard for malicious program. However, implementing CAPTCHAs is becoming increasingly hard due to advances in machine learning system. Furthermore, all current audio-based CAPTCHAs have been broken by automated programs and research shows that the existing implementations are very difficult and time consuming. In addition to this, more than 50 percent of people are unable to bypass the current audio-based CAPTCHAs owing to the intrinsic hardness of interpreting the noisy sound message. Consequently, the implementation of a novel voice CAPTCHA is demanded. In this paper, a technique for telling the human beings and computer programs apart based on submitting the right colour name of the object announced by the speaker has been introduced. In this technique, an object is selected from database at random, and then the selected object will be pronounced from the audio message. After that the CAPTCHA request users to submit the colour name of that object. If the colour name is submitted accurately, the system is able to decide that the client is a human and not an automated program. The main advantage of the proposed CAPTCHA is that users don't have to memorise a group of random digits and words, which stretches the limits of individuals' short-term mind. Finally, the usability test is conducted with some individuals. In addition, discussion, limitations, and suggestions for further study are illustrated.

## Keywords
CAPTCHA, Audio CAPTCHA, Automated program.

## 1. INTRODUCTION
The spam issue is becoming more and more significant as result of the speed global development of online services. For that reason, some companies such as Yahoo and Google now introduced systems in order to deal with malicious programs and spams over online web services [7]. One of the most important difficulties is to identify the types of users when trying to stop spams that derive from automated programs [7]. Alan Turing points out a human tester in "Turing Test" paper in 1950 [18]. The purpose of this test is to differentiate robots from human beings. These days, another way of Turing Test called Reverse Turing Test has been widely used. The word ''Reverse Turing Test'' is employed to classify that the user is not an individual but a robot. In addition to that, this type of test is named CAPTCHA (Completely Automated Public Turing Test to Tell Computer and Humans Apart) in Computer in the field of the web service security [7]. Some internet companies like Google, Yahoo and so on need CAPTCHA to offer services to their customers. On the other hand, the current CAPTCHAs can be broken by some existing software. An audio-based CAPTCHA that appropriate for web systems has been developed in this article. Firstly, background and related study will be shown and the chief features of audio-based CAPTCHAs will be explored. Next, the essential necessities of an audio implementation will be presented and how a voice CAPTCHA is appropriate for online services, then illustrated an algorithm for choosing the proper system.

After that, the process for testing an audio based implementation has been followed. The concept is to present a hard Artificial Problem test with the intention that either the aim of differentiating humans from computer programs is introduced, or that an Artificial Intelligence innovation is accomplished. The security of these mechanisms is not based on the confidentiality of the data field, but on the real hardness of the challenge. The hardness of passing these systems for an automated program and for a legitimate user frequently rises in parallel. Because a CAPTCHA mechanism is infrequently operated independently and frequently included as a supplementary part for systems like web forum, asking for human's concentration for more than a number of seconds is impractical. Therefore, complex problems needing users to spend longer seconds do not make them realistic to be applied on real- world applications.

In order to confuse computer programs, using background noise, recognising distorted alphabets, solving tests based upon pictures or audio files are some methods that are being currently used. On the other hand, with the raising advances in the area of machine learning, automated programs are currently able to identify to defeat image-based and audio based CAPTCHA systems utilizing methods like OCR (Optical Character Recognition) and ASR (Audio Speech Recognition), segmenting audio or image and so on. Raising the difficulty of audio-based and image-based CAPTCHAs by presenting too much noise and distortion in order to make the CAPTCHA harder for computer programs make those implementations harder for human beings as well.

Audio-based schemes have been created to improve the usability of implementations for users with poor vision. Nevertheless, most existing state of the art audio-based schemes like Recaptcha suffer from the shortage of usability. ReCaptcha audio CAPTCHA tests merely concentrate on sound recognition, asking users to recognise all numbers from audio message.

Attackers utilising machine learning methods can succeed a high success rate against audio CAPTCHAs such as Yahoo [11].

A new sound CAPTCHA has been developed in this paper, which has a two-step model that requires the individuals to identify the subject from the audio message and submit the

colour name of that subject. Our audio CAPTCHA is more secure against counterpart machine learning attacks duo to the use of logical correlation. Other audio-based schemes such as Secure Image CAPTCHA [7], Authorize [7] are only available in English. In addition, their usability issue is one of the biggest problems. In this study, further audio CAPTCHA schemes and their drawbacks will be examined in the next sections.

## 2. BACKGROUND

A CAPTCHA is a challenge-response test that most people ought to solve, but bots should not [4]. These kinds of tests usually rely upon difficult open AI difficulties such as automatic identification of distorted text, or of user voice against a sound background [4]. This test is different from the original Turing Test. Because CAPTCHA test are created and evaluated by a machine automatically. As just people are capable of making a rational response, a protocol that has an automated Turing Test can confirm whether the tester is an automated program or a person behindhand the tested computer. Even though, the real Turing Test was invented for evaluating of improvement for AI, CAPTCHA is quite human-nature-verification system [1]. In this paper, the main type of CAPTCHA is audio CAPTCHA. The audio CAPTCHA is designed for people with poor vision as it is difficult for such users to solve text-based CAPTCHA. Thanks to audio CAPTCHA, they can register a service that uses a CAPTCHA. Nowadays, applying visual CAPTCHA in online services is difficult primarily because of the limits of end-user devices. But an audio CAPTCHA is beneficial to protect against automatic audio VoIP messages. For instance, a small number of individuals have a home-based telephony device using a screen that able to show an appropriate (high resolution) image CAPTCHA [1]. If there is a proper CAPTCHA, a spit-bot should not react easily and properly. Therefore, it succeeds to start a call. Moreover, audio CAPTCHA looks reasonable, since text-based CAPTCHA is vulnerable.

CAPTCHAs can mainly be classified into four different groups. Text-based, image-based, audio based and video based [14].

### 2.1 Text Based CAPTCHAs

These types of CAPTCHAs are based on the identification of phrase, alphabets or digits, with a variety of warped applied [15]. Such CAPTCHAs are the earliest to be invented and the oldest to be utilized via OCR (Optical Character Recognition) [15]. Text based CAPTCHAs use various forms utilising different techniques to make the system more difficult against attacks like inserting systematic noise, twisted letters, using multicoloured grounds and overlapping words or digits [15]. Nevertheless, deciphering such CAPTCHAs by a bot typically consists of two steps: i) segmenting character ii) letter identification [15].


**Fig 1. Some popular text based CAPTCHAs [17]**

### 2.2 Image Based CAPTCHAs

This system uses picture classes, which are straightforwardly solved by users but difficult for computers to pass. In order to prevent web services from being exploited, some image based implementations utilize techniques to distort images such as illumination, colour quantization, rotating images, re-scaling etc. [15].

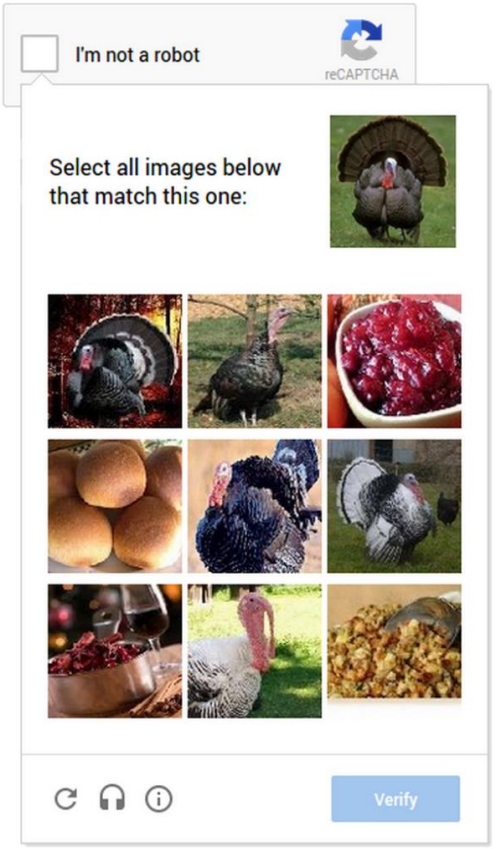


**Fig 2. An example of image based CAPTCHA [20]**

## 2.3 Audio Based CAPTCHAs

These CAPTCHAs are employed as an option for unsighted users as they cannot solve visual-based CAPTCHA. But audio implementations are harder to pass. In this scheme, the users who wish to use web services have to recognise the word or letter pronounced. Audio CAPTCHAs are generally used in the same interface as image-based CAPTCHAs.


**Fig 3. An example of audio based CAPTCHA [19]**

# 3. RELATED WORK

The related research work is now limited duo to the CAPTCHA is almost very new technology and there are a number of important research methodologies that can guide this article.

The authors of [7] developed an audio-based CAPTCHA for VOIP telephony. In addition, they carried out experiments on identifying isolated numbers in the presence of background noise utilising one of the frequency and energy peak detection bots called decaptcha. But their CAPTCHA system is not enough for web services in terms of security duo to the use of a limited data field of digits.

In [12], the authors are able to break eBay audio-based CAPTCHAs with success rate of 75%. They implemented a bot called decaptcha in order to achieve their purpose. An automatic process for downloading audio-based CAPTCHAs is demonstrated to train this decoder, and then passing the eBay audio implementation. This bot can be used to test the security of the proposed audio CAPTCHAs.

The authors of [10] emphasized the usability problems that ought to be taken into account in the development of an audio CAPTCHA. The authors did not particularly work on audio-based tests, with the exception of some properties such as the language used.

In [9], the authors showed that the current audio-based CAPTCHAs are undoubtedly harder and time-consuming to solve in comparison to other types of CAPTCHAs. As well as this, the current audio CAPTCHAs are evaluated in terms of usability but the authors did not demonstrate how the properties of the audio CAPTCHAs influence on the human pass rate. Finally, a new optimized interface for non-visual use that can be used with current audio-based CAPTCHAs created.

The authors of [11] use machine learning techniques to break successfully some popular audio CAPTCHAs such as Google, reCAPTCHA with accuracy up to 71%, which can be followed in this article to test the robustness of the proposed audio CAPTCHAs.

# 4. THE PROPOSED AUDIO-BASED CAPTCHA

In this study, an iteration approach is utilized to implement a novel audio-based CAPTCHA. At first some features that are suitable for the proposed implementation have been chosen. The development of our proposed CAPTCHA is not only reliant upon those features and also reliant upon the cognitive ability of human to solve a challenge. After the design of the system, we will evaluate the system with some humans until the human pass rate is sufficient.

A technique for distinguishing real users from automated programs via recognition and understanding a name of a thing from voice message will be introduced. The framework of the proposed technique is straightforward and shown blow. Initially the system prepares a name of an item at random from the database. The names might be fruits, vegetables, natural things and so on. The proposed implementation selects a name of an item at random and then the CAPTCHA will use some distortion techniques to make the audio message harder for an automated program. After that the system will announce the name of the chosen object. Finally, the system requires the user to enter the colour name of that object accurately. For instance, the CAPTCHA selects “snow” as an objective and then asks the tester to enter the right colour name of that pronounced object into the textbox. If the tester submits the right colour name of that object, the system decides that the tester is not automated program is a human. Since an automated program ought to correctly make these successive actions to react properly. At first the automated program has to identify the name of the pronounced item and then know the colour of that item. After that it has to submit the name of this colour entering this name into the text box and pressing the submit button.

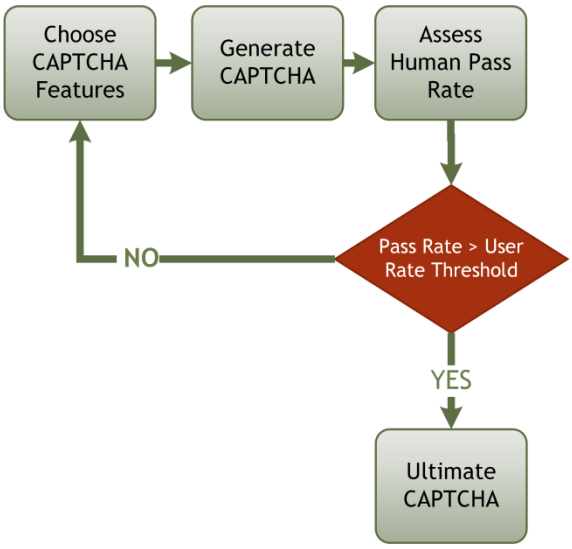


**Fig 4. The development process of the audio-based CAPTCHA**

## 4.1 Usability and Security Testing

The test in answering the proposed implementation consists of two steps. At first the tester must recognize the name of the pronounced object from the distorted audio message and then decide the colour of the selected object and submit this colour name into the textbox. The proposed CAPTCHA resolving capability comes to human beings in nature since users utilise the human cognition skill and good sense without even the intrinsic hardness of the test. The identical test for an automated program can ask both identifying the object announced from voice message and finding the related colour of that object, generating a hard Artificial Intelligence difficulty. This proposed scheme intends to raise the hardness for automated software and make this CAPTCHA more

usable for users, without sacrificing the security of the implementation.

The concept of introducing humans with two-step problem "recognising the pronounced object from the audio message and submitting the colour name of that object" would make this CAPTCHA novel. The thought of identifying the colour name of the pronounced object has wider scope than common recognition of a group of numbers or letters such as recognising numbers in Yahoo CAPTCHA. Such aspect makes our system different from current audio-based systems that just ask individuals to decipher the initial step, which is audio identification. It is difficult for an automated program to understand and recognise the colour name of the announced object, which makes this CAPTCHA more secure against malicious programs.

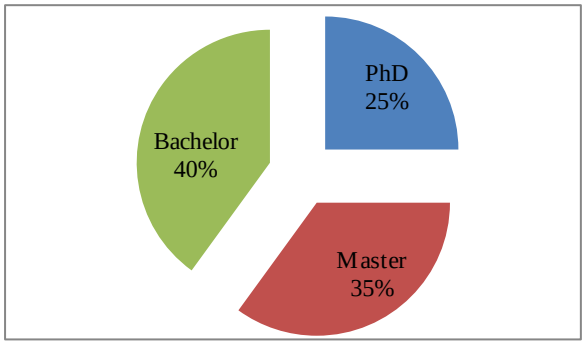


**Fig 5. Education level of users**

Our usability study is conducted in which some individuals partake. The educational levels of participants are Bachelor, Master and PhD as illustrated in Fig 5. The aim of this test is to evaluate the system in terms of usability.

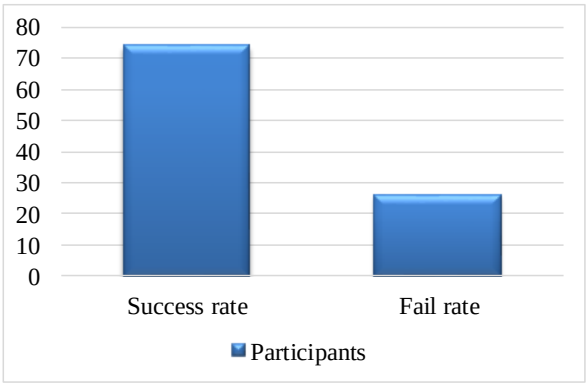


**Fig 6. The user success rate**

The usability study is performed as follows:

1. The user is required listening the audio CAPTCHA shownon a desktop in person and to solve what they understand. Users have 6 seconds to complete the task. If the user is unable to solve the challenge, then the second chance is given to the user.
2. After announcing the audio message, the user will answer the test, which what the user think to have heard.
3. The user's answer is calculated, correct or incorrect, all the time.

The results of the test are that about 90 percent of the users pointed out that they found the proposed system is simple, at the same time 35 percent of the users said that the CAPTCHA is straightforward. 83 percent of users pointed out that this CAPTCHA is entertaining, but 27 percent of the users found it more entertaining. About 80 percent of the users were pleased, but about 20 percent of the users were dissatisfied when trying the challenge.

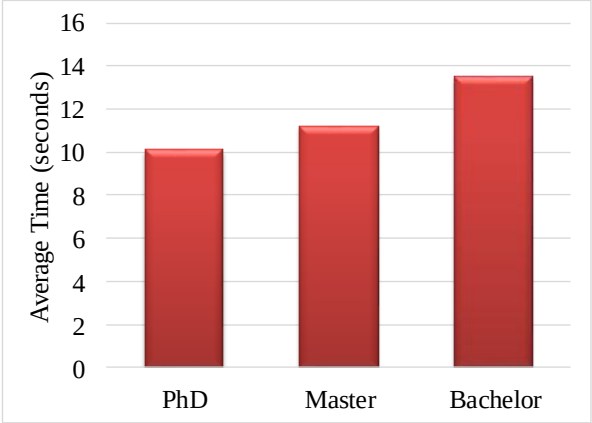


**Fig 7. The average time for solving CAPTCHA for each different group**

The approximate time to pass the system is about 11.5 seconds, but the approximate time to pass final version with distortion is 13.6 seconds. Finally the approximate pass rate is %74, which is high enough compared to current audio CAPTCHAs as illustrated in **Error! Reference source not found.**.

## 5. CONCLUSIONS

In this study, a technique for telling humans and automated software apart in the case of using both the human cognitive ability and distorted audio has been introduced. English language is the most common verbal communication all over the world. Consequently, our CAPTCHA supports a wide range of humans on online.

It is observed that differentiating users from automated programs; CAPTCHA provides excellent defence against automatic exploitation on web services and online applications. The criteria for success of an audio CAPTCHA are its security and usability. In addition, some techniques are presented to create the challenge that renders the proposed audio CAPTCHA secure since it is robust against malicious software that use ASR (Automatic Speech Recognition) technology.

Our framework is very usable since it is straightforward for users to correctly pass the challenge. Therefore, the pass rate of the users is high. There are still some chances for development and improvement of those techniques, which needs more study to be completed.

## 6. REFERENCES

[1] Gao, H., Liu, H., Yao, D., Liu, X. and Aickelin, U., "An audio CAPTCHA to distinguish humans from computers," in 3rd International Symposium on Electronic Commerce and Security, ISECS 2010, 2010, pp. 265–269.

[2] Bursztein, E., Beauxis, R.,Paskov, H.,Perito, D., Fabry, C. and Mitchell, J., "The failure of noise-based non-continuous audio CAPTCHAs," in Security and Privacy (SP), 2011 IEEE Symposium on, 2011, pp. 19-31.

[3] Almazyad, A. S., Ahmad, Y. and Kouchay, S. A. , "Multi-modal captcha: A user verification scheme," in Information Science and Applications (ICISA), 2011 International Conference on, 2011, pp. 1-7.

[4] Rahman, R. U., Tomar, D. S. and Das, S. "Dynamic Image Based CAPTCHA," in Communication Systems and Network Technologies (CSNT), 2012 International Conference on, 2012, pp. 90-94.

[5] Chan, T.-Y., "Using a test-to-speech synthesizer to generate a reverse Turing test," in Tools with Artificial Intelligence, 2003. Proceedings. 15th IEEE International Conference on, 2003, pp. 226-232.

[6] Baird, H. S., Coates, A. L. and Fateman, R. J., "PessimalPrint: a reverse Turing test," International Journal on Document Analysis and Recognition, vol. 5, pp. 158-163, 2003.

[7] Soupionis, Y. and Gritzalis, D., "Audio CAPTCHA: Existing solutions assessment and a new implementation for VoIP telephony," Computers & Security, vol. 29, pp. 603-618, 2010.

[8] Szczys, M., (2012, 30 August 2015). Stiltwalker beat audio reCAPTCHA. Available: http://hackaday.com/2012/06/15/stiltwalker-beat-audio-recaptcha/.

[9] Bigham, J. P. and Cavender, A. C., "Evaluating existing audio CAPTCHAs and an interface optimized for non-visual use," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2009, pp. 1829-1838.

[10] Lee, Y.-L. and Hsu, C.-H., "Usability study of text-based CAPTCHAs," Displays, vol. 32, pp. 81-86, 2011.

[11] Tam, J., Simsa, J., Hyde, S., & Ahn, L. V. (2008). Breaking audio captchas. In Advances in Neural Information Processing Systems (pp. 1625-1632).

[12] Tam, J., Simsa, J., Hyde, S., and Ahn, L. V., "Breaking audio captchas," in Advances in Neural Information Processing Systems, 2008, pp. 1625-1632.

[13] C. Technologies. (2010, 20 August 2015). Image-Based CAPTCHA for PHP 1.5. Available: http://www.phpkode.com/scripts/item/image-based-captcha-for-php/.

[14] Kluever, K. A. and Zanibbi, R., "Balancing usability and security in a video CAPTCHA," in Proceedings of the 5th Symposium on Usable Privacy and Security, 2009, p. 14.

[15] Obimbo, C., Halligan, A. and De Freitas, P., "CaptchAll: an improvement on the modern text-based CAPTCHA," Procedia Computer Science, vol. 20, pp. 496-501, 2013.

[16] Bursztein, E., Martin, M. and Mitchell, J., "Text-based CAPTCHA strengths and weaknesses," in Proceedings of the 18th ACM conference on Computer and communications security, 2011, pp. 125-138.

[17] http://www.gsacaptchabreaker.com/wp-content/uploads/2013/02/captcha-examples.png. Last accessed 6 December 2015.

[18] Turing, A. M., "Computing machinery and intelligence," Mind, pp. 433-460, 1950.

[19] S. McGlaun, 'Google Street View and reCAPTCHA Get Smarter with New Algorithm', dailytech.com, 2015. [Online]. Available: http://www.dailytech.com/Google+Street+View+and+reCAPTCHA+Get+Smarter+with+New+Algorithm/article34740.htm. [Accessed: 07- Dec- 2015].

[20] Available:http://i.kinjaimg.com/gawkermedia/image/upload/t_original/c8ayzgvhgvaasa8kwhzo.png. [Accessed: 07- Dec- 2015].