マウスを用いた単純操作の音声型CAPTCHAの提案

藤　竜成 [a)]・坂本 竜也 [b)]・山場　久昭 [c)]・油田　健太郎 [d)]・岡崎　直宣 [e)]

Proposal of Audio Based CAPTCHA with Simple Operation Using Mouse

Ryusei FUJI, Tatsuya SAKAMOTO, Hisaaki YAMABA, Kentaro ABURADA, Naonobu OKAZAKI

Abstract

In the Internet, bots which send large amounts of spam comments on bulletin boards, blogs, etc., and obtain webmail service accounts illegally hinder the smooth operation of the web service. In order to prevent the above bot activeties, a system named CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) is used. Typical examples of CAPTCHA include a character string based CAPTCHA that users need to interpret distorted character strings, and an image based CAPTCHA that users need to select images conforming to questions from multiple images. However, these methods are hard to decipher for visually impaired people. Therefore, we adopt another CAPTCHA called audio based CAPTCHA. Among audio based CAPTCHAs, a speech recognition based CAPTCHA is a main method, but this method has problems such as difficulty in listening to the speech, time required for answering, low attack resistance by bots. In this paper, we propose an audio based CAPTCHA which can respond to visually impaired people's use and solve the CAPTCHA in a short time with a simple operation. In the proposed method, the user need to perform a mouse operation when a specific environmental sound is heard, and from the operation result, it is determined whether the user is a human or a bot. In addition, we implemented the proposed method CAPTCHA and conducted an experiment to evaluate practicality and usability. As a result of the experiment, we archived a high correct answer rate, and the SUS score, which is an index of the usability evaluation, also reached a high score.

Keywords: CAPTCHA, audio based, accessibiity, environmental sound

1. はじめに

インターネットにおいてボットが掲示板、ブログなどにおいてスパムコメントを大量に送信したり、ウェブメールサービスのアカウントを取得して大量の迷惑メールを送信したりすることによって、サービスの円滑な運営が妨げられている。これを防止するために CAPTCHA（Completely Automated Public Turing test to tell Computers and Humans Apart）という対象者が人間かボットかを判別するシステムが用いられる。

CAPTCHA の代表例として、歪んだ文字列を解釈させる文字列型 CAPTCHA(図 1) や複数の画像から出題に沿った画像を選択する画像型 CAPTCHA(図 2) が挙げられるが、これらの方式は視覚障がい者には解読困難である。そのため、音声を利用した音声型 CAPTCHA がある。音声型 CAPTCHA の中でも、英数字識別型 CAPTCHA(図 3) が主流な方式であるが、この方式には、音声の聞き取りが難しい、解答に時間がかかる、ボットによる攻撃耐性が低いなどの問題点がある [1)]。

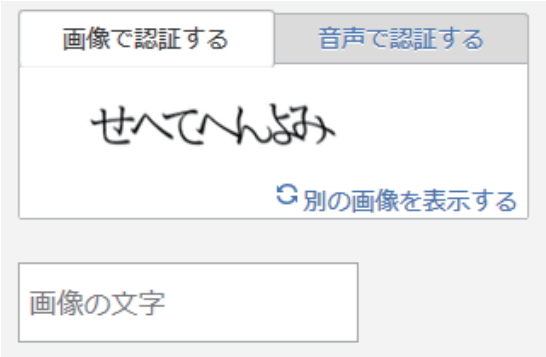

図 1. 文字列型 CAPTCHA の例

そこで本論文では、視覚障がい者の利用に対応し単純な操作により短時間での解答を行える音声型 CAPTCHA を提案する。提案方式では、聴力検査を模し、特定の環境音が聞こえている間マウス操作を行い、その結果から人間か機械かを判別する。

2. 先行研究

2.1 既存の音声型 CAPTCHA

既存の音声型 CAPTCHA の主流である、英数字識別型 CAPTCHA は、ランダムな英数字を音声で流し、対象者に聞き取った内容を解答させる方式である。ボットによる解答

[a)] 工学専攻機械・情報系コース大学院生
[b)] 情報システム工学科学部生
[c)] 情報システム工学科助教
[d)] 情報システム工学科准教授
[e)] 情報システム工学科教授

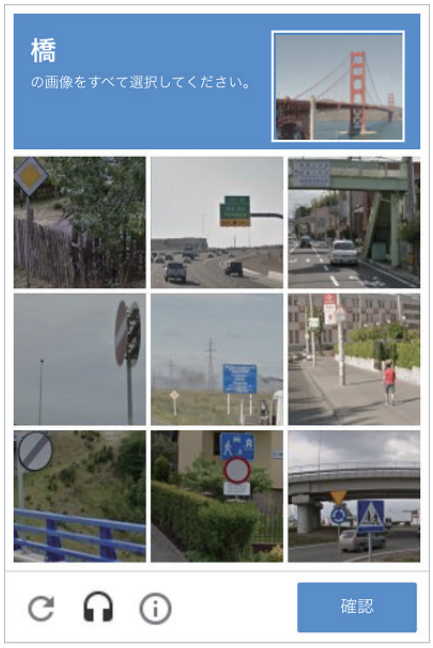

図 2. 画像型 CAPTCHA の例

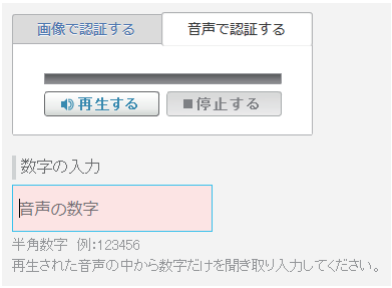

図 3. 音声型 CAPTCHA の例

を困難にするため、音声に歪みを加えたり、英数字以外の音声が流れるなどの工夫がなされている。この方式は Yahoo! JAPAN[2] など、様々なウェブサイトで利用されている。

2.2 既存の音声型 CAPTCHA の問題点

既存の音声型 CAPTCHA には以下のような問題点が存在する [1]。

(1) 音声の聞き取りが難しい

音声認識技術の向上に対抗し、さらに音声の歪みを大きくすることにより、英数字識別型 CAPTCHA は人間にとっても困難なものが存在する。

(2) 解答に時間がかかる

音声型 CAPTCHA は正答率が低く、認証されるまで何度も問題を解かなければならない場合がある。

また、視覚障がい者らが使用するスクリーンリーダでは、キーボードの入力結果を読み上げるため、音声の聞き取り中に解答すると自らの操作音が音声認識の邪魔になる [3]。

そのため、視覚障がい者は音声を聞き取りながら入力をする作業が困難であるとされ、聞き取った音声を記憶する必要がある英数字識別型 CAPTCHA は大きな負担となる。

(3) ボットによる攻撃に耐性が低い

英数字文字列型 CAPTCHA に対して、機械学習による攻撃成功例があり、Google reCAPTCHA に対し 58.75%[4]、Yahoo!、Microsoft、eBay の 3 つのウェブサイトの CAPTCHA に対して、45%、49%、83%の確率で攻撃に成功している [5]。このことから、機械学習により歪みの加わった音声であっても識別することが可能であり、ボットによる攻撃への耐性が弱いと考えられる。

3. 提案手法

3.1 提案 CAPTCHA

提案 CAPTCHA では、2 種類の環境音が流れる。1 つが CAPTCHA 開始時から終了時まで流れるノイズとしての役割を持つ背景音であり、もう 1 つが判定に使用する目標音である。

背景音として、日常生活において数十秒以上持続して聴こえる種類の長時間の音、目標音として、日常生活において音の発生から消滅までが数秒以内の種類の短時間の音を使用する。

この目標音が流れている間、ユーザーが指示通りにマウス操作を行うことで正誤判定が行われ、正しいタイミングで操作を実行できていれば人間とみなす。

環境音を使用する理由として、普段から誰もが聞き慣れていることにより、音の識別や音への反応が容易である点と、種類の膨大さにより、機械学習を用いたボットによる攻撃に耐性があると期待できる点である。

3.2 認証手順

提案する CAPTCHA を用いた認証手順を図 4 に示す。CAPTCHA プログラムが動作を始めると、「○○の音が聞こえている間、マウスを長押ししてください。」などの音声案内が再生される。その 1 秒後、背景音が 10 秒間が再生される。この 10 秒間のうちのランダムな箇所で目標音が再生される。ただし、背景音と同時に開始、または終了することはない。この目標音が開始時されて 0.7 秒以内にマウスを押し、終了してから前後 0.7 秒以内にマウスを離すことが出来た場合のみ、認証成功とする。マウスを押す、離す操作はそれぞれ 1 度のみ行える。

3.3 実装

提案 CAPTCHA の開発言語は JavaScript である。音声案内の作成は、SofTalk[6] という合成音声によるフリーのテキスト読み上げソフトを利用している。

3.4 判定について

感覚刺激の提示から行動による反応が生じるまでに経過した時間のことを反応時間という。聴覚刺激の検出における反応時間は 0.14 秒から 0.16 秒である [7]。

しかし、使用する PC の環境が影響したためか、事前実験では平均反応時間は約 0.4 秒、遅いもので約 0.6 秒となっており、報告とは異なる結果となった。具体的に、事前実験では、宮崎大学に所属する 20 代の健常な学生 5 名に対し、音が聴こえてから何秒以内にマウス操作が行えるかという事前実験を 1 人あたり 5 回、合計 25 回行った。これは反応時間が短くなる 20 代の結果であり、年齢の増加とともに長くなる可能性が考えられる。

そこで、提案 CAPTCHA は目標音が聴こえ始めてから 0.7 秒以内、聴こえ終わってから前後 0.7 秒以内にマウス操作を

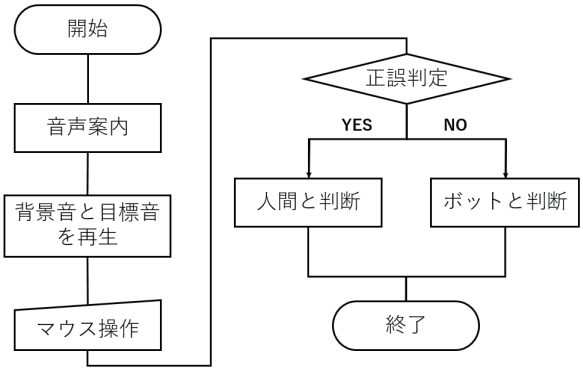

図 4. 認証手順のフローチャート

行えればよいものとする。判定において、正答とするタイミングを赤枠で示したものを図 5 に示す。

3.5 安全性

提案方式に対する、ボットによる攻撃手法として、以下の手順を考える。

(1) 音声案内を解析し、どの環境音が目標音に設定されたのかを理解する。
(2) 背景音と目標音が合成された環境音を解析し、それぞれの音に分離する。
(3) 分離された音がそれぞれ何を表す音なのかを解析し、目標音がどちらかを理解する。
(4) 目標音が開始されて 0.7 秒以内にマウスを押す、終了してから前後 0.7 秒以内にマウスを離すという処理を行う。

以上の手順 (2) と (3) に対して、それぞれ独立成分分析、機械学習を用いた攻撃が考えられる。この中で、手順 (3) に対する機械学習を用いた音響イベントの識別が最も困難だと考えられている [1)]。

音響シーンと音響イベントの識別精度を競うコンテスト DCASE2018 Challenge では、様々な環境音（人の声、犬や猫の鳴き声、アラーム、掃除機、髭剃りなど）が録音された一定時間の音声データを元に、どの時間帯にどの種類の音がしたかをプログラムで識別する課題が提示された。その課題の結果である識別率の F スコアは、最も識別率の高い手法であっても、32.4%であった。

このように、環境音の音響イベント識別はまだ難しいということが分かる。さらに、手順（2）への独立成分分析による攻撃に成功しなければならないため、提案 CAPTCHA はボットによる攻撃に耐性があると期待できる。

4. 評価実験

4.1 実験目的

今回、提案 CAPTCHA はユーザー（人間）にとって解答可能なものであるかを検証するため、ユーザーの成功率を調査し、更にユーザビリティ評価を行うことで CAPTCHA としての実用性について調査する。

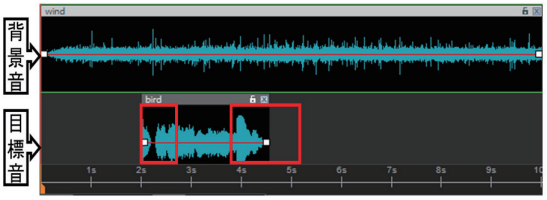

図 5. 判定の視覚化

4.2 実験方法

被験者は宮崎大学に所属する 20 代の健常な学生 20 名である。被験者には提案 CAPTCHA を解いてもらう前に一通りの操作手順を説明し、本人が慣れたと感じるまで最低 1 度の練習を行ってもらった。被験者には、背景音 3 種、目標音 5 種の組み合わせ 15 パターンの問題を 1 パターン 1 度ずつ解いてもらい、被験者が操作を行った結果である各回答の正否とページ表示から解答までの所要時間を記録した。使用した環境音の組み合わせの詳細を表 1 に示す。

またユーザビリティ評価を行うために、各被験者の正答率や所要時間の測定を行った後、アンケート調査を行った。アンケート調査では、 SUS(System Usability Scale) というユーザビリティの数値的な評価が可能である指標を用いた。SUS についての簡単な説明とアンケートの質問項目を 4.4 に示す。

4.3 環境音の選択

本論文では環境音の収集を簡易にするため、フリー素材として効果音などを配布するサイト [8)] を利用した。選択の基準として、背景音の長さは 10 秒以上であること、目標音の長さは 1 秒から 2 秒のみを設定した。

結果として、背景音は、雨、街の道路、風の 3 種類、目標音は、鐘、鶏の鳴き声、ドアを叩く音、ガラスの割れる音、山羊の鳴き声の 5 種類を選択した。

4.4 SUS(System Usability Scale)

SUS(System Usability Scale)[9)] は John Brooke が 1868 年に開発したもので、ユーザビリティの評価のために多く利用されている質問票である。以下は、評価に用いられる 10 項目のアンケート内容である。

(1) この CAPTCHA をしばしば利用したいと思う。
(2) この CAPTCHA は必要以上に複雑であると感じた。
(3) この CAPTCHA は容易に使いこなすことができると思った。
(4) この CAPTCHA を利用するのに専門家のサポートが必要だと感じる。
(5) この CAPTCHA が提供する様々な機能は統一性があると感じた。
(6) この CAPTCHA には統一性のない部分が多々あったと感じた。
(7) この CAPTCHA は大半の人がすぐに使用方法を理解すると思った。
(8) この CAPTCHA はとても操作しづらいと感じた。
(9) この CAPTCHA を利用できる自信がある。
(10) この CAPTCHA を利用するために多くのことを学ばなければならないと感じた。

4.4.1 SUS の集計方法

各項目は 1〜5 で評価され、その後以下のプロセスを経て SUS を計算する。
奇数項目：回答番号から 1 を引く
偶数項目：5 から回答番号を引く
すべての項目は 0 から 4 で評価し、足しあわせた合計数値を 2.5 倍して 0 から 100 のスケールへ変換する。

各項目のスコアを N_1 から N_{10} とすると、合計スコア S は式 (1) で表すことができる。

$$S = (\sum_{i=1}^{10} N_i) \times 2.5 \quad (1)$$

スケール後の数値が高いほど、システムとして良い評価が与えられる。SUS スコアは、Jeff Sauro らによる調査結果 [9] から平均スコアが 68 とされており、ユーザビリティに優れた上位 10%に入るには、80.3 を超えるスコアが必要とされている。

4.5 実験環境

本実験の環境は以下のとおりである。

OS：Windows 10(64bit)
CPU：Intel(R) Core(TM) i7-7700 CPU @ 3.60GHz
メモリ：16.0GB

4.6 実験結果と考察

実験結果から、全てのパターンの正答率は 85%以上、解答時間は 6 秒〜10 秒である (表 2)。この結果に対して、既存の音声型 CAPTCHA の平均正答率は 31.2%、平均解答時間は 28.4 秒であり [10]、提案 CAPTCHA は既存の音声型 CAPTCHA より正答率が高く、短時間で解答できる CAPTCHA であると考えられる。

また今回の SUS に基づいたアンケート調査のスコアは 85 であり、優れたユーザビリティを示すスコア 80.3 を超える結果となった。

これらのことから、 提案 CAPTCHA は実用的であるといえる。

また、背景音と目標音をそれぞれ、数十秒以上持続する長時間の音と音の発生から消滅までが数秒以内の短時間の音に区別することにより、人間にとって目標音を聞き取りやすくなったと考えられる。ただし、背景音として用いた音が長時間持続するような音、目標音が比較的短時間の音だという人間の知識が聞き取りやすさに貢献したのか、単に背景音と目標音が異なる音だというだけで判別できていたのかどうかは現段階では不明である。

問題点として、ガラスの割れる音は対象者に不快感を与える可能性があるという意見が挙げられたため、

今後は、環境音の中を無作為に選択せず基準を設けて選別するなど、さらに人間にとっては使いやすい手法を検討する必要がある。

表 1. パターン詳細

パターン	背景音	目標音
パターン 1	雨	鐘
パターン 2	雨	鶏の鳴き声
パターン 3	雨	ドアを叩く音
パターン 4	雨	ガラスの割れる音
パターン 5	雨	山羊の鳴き声
パターン 6	街の道路	鐘
パターン 7	街の道路	鶏の鳴き声
パターン 8	街の道路	ドアを叩く音
パターン 9	街の道路	ガラスの割れる音
パターン 10	街の道路	山羊の鳴き声
パターン 11	風	鐘
パターン 12	風	鶏の鳴き声
パターン 13	風	ドアを叩く音
パターン 14	風	ガラスの割れる音
パターン 15	風	山羊の鳴き声

表 2. 実験結果

パターン	平均解答時間 [秒]	正答率 [%]
パターン 1	8.23	100
パターン 2	8.67	90
パターン 3	8.46	90
パターン 4	8.24	100
パターン 5	8.27	85
パターン 6	7.95	90
パターン 7	8.97	95
パターン 8	8.63	95
パターン 9	8.95	100
パターン 10	8.35	100
パターン 11	7.63	95
パターン 12	9.66	100
パターン 13	7.93	100
パターン 14	8.98	100
パターン 15	6.64	95

5. まとめ

本研究では、既存の音声型 CAPTCHA の音声の聞き取りが難しい、解答に時間がかかる、ボットによる攻撃耐性が低いなどの問題点に着目し、視覚障がい者のアクセシビリティを損なわない、単純な操作により短時間での解答を行う音声型 CAPTCHA を提案した。また、提案方式の CAPTCHA を実装し、実用性とユーザビリティ評価を行う実験を行った。実験の結果、人間であるユーザーが解いた場合は高い正答率を示し、ユーザビリティ評価の指標となる SUS スコアも高い数値となった。これらのことから、提案 CAPTCHA が実用的であることが確認できた。

今後は、今回確認することができなかった新たな攻撃方法による自動プログラムへの耐性を検証しつつ、提案 CAPTCHA により適した環境音の調査や、環境音の選択に基準を設けたことにより、簡易に収集できなくなるため、収集法の最適化などの問題点の解決に向けて検討していかなければならない。

参考文献

1) 古賀 千裕, 佐藤 敬: 混合された環境音の聞き取りに基づく認証方式, コンピュータセキュリティシンポジウム 2017 論文集,Vol.

2017,No.2,2017.

2) https://www.yahoo-help.jp/app/answers/detail/p/533/a_id/87423/~/画像や音声による認証について, (2019/01/28 閲覧).

3) 山口 通智, 菊池 浩明: 多様な話者により発話されたランダムな音韻列と単語の識別問題を用いた音声型 CAPTCHA の研究, コンピュータセキュリティシンポジウム 2016 論文集, Vol.2016, No.2, (2017) ,pp.363-370, 2008.

4) S.Sano, T.Otsuka, K.Itoyama, and Hiroshi G. Okuno: HMM-based Attacks on Google's ReCAPTCHA with Continuous Visual and Audio Symbols, In *Journal of information processing*, Vol.23, No.6 pp.814-826, 2015.

5) E.Bursztein, R.Beauxis, H.Paskov, D.Perito, C.Fabry, and J.Mitchell: The Failure of Noise-Based Non-Continuous Audio Captchas, In *2011 IEEE Symposium on Security and Privacy*, pp.19-31, 2011.

6) https://www35.atwiki.jp/softalk/, (2019/01/28 閲覧).

7) I.Muhammad: The Spirit of Muslim Culture, In *The Reconstruction of Religious Thought in Islam*, (2019/01/28 閲覧).

8) https://soundeffect-lab.info/, (2019/01/28 閲覧).

9) J.Sauro, MEASURING USABILITY WITH THE SYSTEM USABILITY SCALE (SUS), Measuring U, `https://measuringu.com/sus/`, (2019/01/28 閲覧).

10) E.Bursztein, S.Bethard, C.Fabry, John C. Mitchell, and D.Jurafskyl: How Good are Humans at Solving CAPTCHAs? A Large Scale Evaluation, In *IEEE Computer Society Washington*, pp.399-413, 2010.