

環境音を利用したマウスによる音声型 CAPTCHA の提案

本田 佳鈴^{a)}・岩村 隆晴^{b)}・油田 健太郎^{c)}・山場 久昭^{d)}・岡崎 直宣^{e)}

Proposal of Audio Based CAPTCHA with Mouse Using Environmental Sound

Karin HONDA, Ryusei IWAMURA, Kentaro ABURADA, Hisaaki YAMABA, Naonobu OKAZAKI

Abstract

The smooth operation of the service is hindered by the bot sending a large amount of spam comments on bulletin boards, blogs, etc. on the Internet, or acquiring a web mail service account and sending a large amount of unsolicited e-mail. In order to prevent this, a system called CAPTCHA that determines whether the target person is a human or a bot is used. However, character-type CAPTCHA and image-type CAPTCHA, which are widely used, are difficult for visually impaired people to decipher. Therefore, in the previous research, a voice-based CAPTCHA was proposed that can respond to the use of the visually impaired in a short time by simple operation. However, in the CAPTCHA of the previous research, the attack using the independent component analysis by the bot was assumed, and it was necessary to consider a new method. In this paper, in consideration of bot attacks against CAPTCHA in previous research, we proposed a new CAPTCHA method with bot resistance and conducted experiments to evaluate its practicality and usability. As a result, it was proved that the proposed CAPTCHA is a CAPTCHA that can be answered in a short time, and that it is a CAPTCHA that is easy for humans to use.

Keywords: CAPTCHA, audio based, accessibility, 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 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 1) 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 2) 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⁾。

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

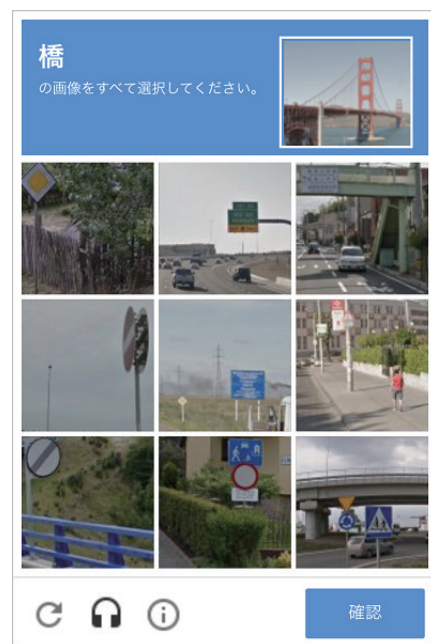


図 1. 画像型 CAPTCHA の例 [1]

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。

^{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准教授

^{d)}情報システム工学科助教

^{e)}情報システム工学科教授

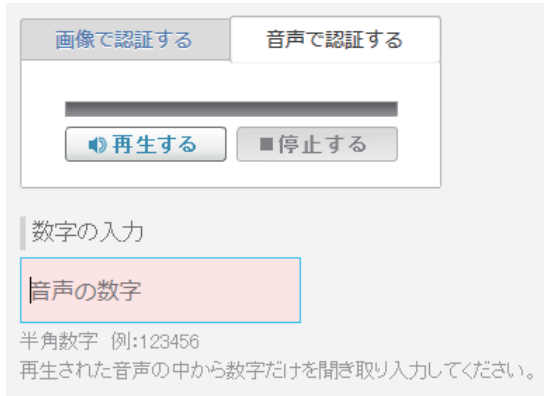


図 2. 既存の音声型 CAPTCHA の例 [1]

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! JAPAN²⁾ など、様々なウェブサイトで利用されている。

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。

2.2 マウスを用いた単純操作の音声型 CAPTCHA

先行研究において、坂本ら⁶⁾ は、2種類の環境音が流れる音声型 CAPTCHA (図 3) を提案した。1つ目の環境音は、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る。2つ目の環境音は、判定に使用する目標音である。この目標音が流れている間、ユーザーはマウスを

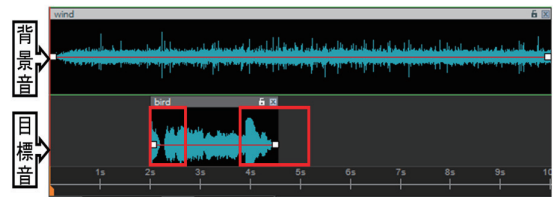


図 3. マウスを用いた単純操作の音声型 CAPTCHA [6]

長押しし続ける。この時の反応時間が前後 0.7 秒以内で操作を実行できた場合、人間とみなす。

環境音を使用した理由は、機械による環境音のシーン識別は困難であると考えたからである。音響によるシーン識別などの精度を競うコンテスト DCASE2019 Challenge⁷⁾ では、音のデータセットを 15 個のクラスタに分類した時の平均精度は 85.2%であった。しかし、公園や電車など特定のシーンにおいては 13.9%、33.6%とあまり高い精度が得られていない、このように、環境音のシーン識別はまだ難しいということが分かる。

坂本らの CAPTCHA 方式では、環境音が 2 種類しか使用されておらず、背景音と目標音の波長の違いにより、機械による独立成分分析を用いた攻撃が考えられる¹⁾。そのため、ボットによる攻撃を考慮した新しい CAPTCHA 方式を検討する必要があると考える。

3. 提案手法

3.1 提案 CAPTCHA

本研究では、独立成分分析への対策として、環境音の種類を増やし、ボット耐性を持たせた新たな CAPTCHA 方式を作成することを目的とする。

提案 CAPTCHA では、5 種類の環境音が流れ、その中で音声案内で指定された 1 つの環境音だけが判定に使用する目標音である。

環境音は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短時間の音を使用する。

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。

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。」と音声案内が再生される。その 1 秒後、環境音が 5 種類再生される。この 5 種類の環境音のうち、ランダムな順番で目標音が再生される。ただし、環境音 1 つ 1 つの間には約 1 秒ほど無音の間隔がある。この目標音が開始されて 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。

提案 CAPTCHA の例として、5 種類の環境音の内 3 番目に目標音がある場合、判定において正答とするタイミングを赤棒で示したものを図 5 に示す。

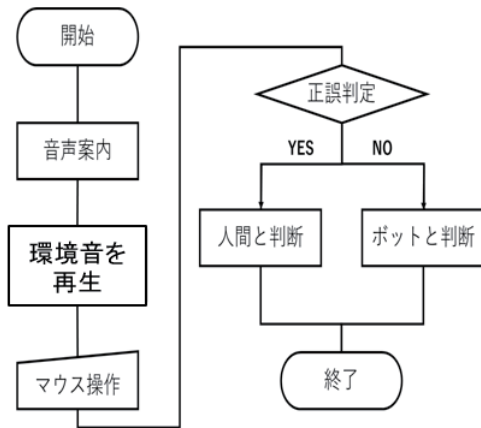


図 4. 認証手順のフローチャート

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、SofTalk⁸⁾ 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⁹⁾。

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 代の健全な学生 10 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 3 回、合計 30 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行えばよいものとする。

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. 5 種類の環境音の中から目標音がどれであることを判別する。
3. 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) に対して、機械学習を用いた攻撃が考えられる。機械による音響シーン識別は 2.2 で前述したとおり、困難だと考えられている。

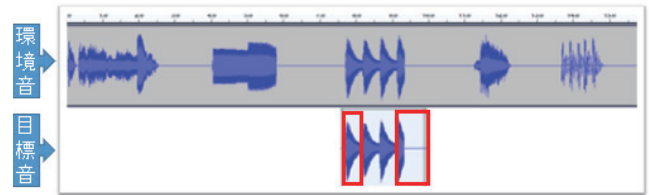


図 5. 提案 CAPTCHA の正誤判定

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。

4.2 実験方法

被験者は宮崎大学に所属する 20 代の健全な学生 10 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、使用する環境音 10 種類が 1 回ずつ目標音になるように問題を 10 パターン解いてもらい、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¹⁰⁾ を利用した。選択の基準として、環境音の長さは 1 秒から 2 秒のみを設定した。

環境音は、鐘、鶏の鳴き声、ドアを叩く音、カラスの鳴き声、ヤギの鳴き声、学校のチャイム、踏切、雨、セミの鳴き声、携帯の着信音の 10 種類を選択した。

4.4 SUS(System Usability Scale)

SUS(System Usability Scale)¹¹⁾ 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。

表 1. パターン詳細と実験結果

パターン	環境音 1	環境音 2	環境音 3	環境音 4	環境音 5	目標音	解答時間 [秒]	正答率 [%]
1	鐘	鶏	ドア	雨	踏切	踏切	20.41	100
2	セミ	カラス	ドア	チャイム	踏切	チャイム	15.73	80
3	踏切	ヤギ	セミ	カラス	携帯	セミ	11.03	70
4	踏切	カラス	セミ	ドア	携帯	カラス	8.49	30
5	ドア	ヤギ	セミ	カラス	携帯	ドア	7.19	60
6	チャイム	鐘	セミ	カラス	携帯	鐘	7.47	60
7	踏切	チャイム	セミ	鶏	鐘	鶏	16.23	80
8	踏切	鐘	ヤギ	カラス	携帯	ヤギ	11.14	70
9	踏切	ヤギ	セミ	カラス	携帯	携帯	18.89	90
10	雨	ヤギ	セミ	カラス	携帯	雨	5.91	40

- この CAPTCHA には統一性のない部分が多々あったと感じた。
- 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。
- この CAPTCHA はとても操作しづらいと感じた。
- この CAPTCHA を利用できる自信がある。
- 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。

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

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¹¹⁾から平均スコアが 68 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、全てのパターンの平均正答率は 68%、平均解答時間は 12.25 秒である (表 1)。

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¹²⁾、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。

今回の SUS に基づいたアンケート調査のスコアは 69 であり、平均スコアの 68 を超える結果となった。

また実験結果からは、環境音の種類によって正答率に差が出たと考えられる。パターン 1 の踏切の音は正答率 100% に対し、パターン 4 のカラスの鳴き声は正答率が 30% と低いことがわかった。

さらに、環境音が流れる順番に対しても正答率に差が出たと考えられる。目標音が流れる順番が前半 (1、2 番目) では 30%~60% と正答率が低く、後半 (4、5 番目) では 80%~100% と高いことがわかった。

アンケート調査から、音声案内の機械音が聞き取りづらく、指定された目標音が何かわからないという指摘があった。普段から機械音を聞き慣れている人とそうでない人の間で正答率に差が出たと考えられる。

今後は、音声案内も人間に聞き取りやすい手法を検討する必要がある。

5. まとめ

本研究では、先行研究の CAPTCHA に対してのボットによる攻撃を考慮し、ボット耐性を持たせた新たな CAPTCHA 方式を提案した。

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、既存の音声型 CAPTCHA と比べて高い正答率を示し、ユーザビリティ評価の指標となる SUS スコアも平均を超える数値となった。

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、人間が聞き取りやすい音声案内の手法を模索していかなければならない。

参考文献

- 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, pp.23-25,2017.

- 2) https://www.yahoo-help.jp/app/answers/detail/p/544/a_id/41993, (2020/2/12 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, pp.363-370, 2016.
- 4) S.Sano, T.Otsuka, K.Itoyama and Hiroshi G. Okuno: HMM-based Attacks on Google's ReCAPTCHA with Continuous Visual and Audio Symbols, Journal of information processing, Vol.23, 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, 2011 IEEE Symposium on Security and Privacy, pp.19-31, 2011.
- 6) 坂本 竜也, 山場 久昭, 油田 健太郎, 岡崎 直宜: マウスを用いた単純操作の音声型 CAPTCHA の提案, 宮崎大学工学部紀要, pp.239-243, 2019.
- 7) <http://dcase.community/challenge2019/>, (2020/2/12 閲覧)
- 8) <https://www35.atwiki.jp/softalk/>, (2020/2/12 閲覧)
- 9) I.Muhammad: The Reconstruction of Religious Thought in Islam, Stanford Univ Pr , 2013.
- 10) <https://soundeffect-lab.info/>, (2020/2/12 閲覧)
- 11) <https://measuringu.com/sus/>, (2020/2/12 閲覧).
- 12) E.Bursztein, S.Bethard, C.Fabry, John C. Mitchell and D.Jurafsky: How Good are Humans at Solving CAPTCHAs? A Large Scale Evaluation, IEEE Computer Society Washington, pp.399-413, 2010.