

オノマトペ CAPTCHA の開発と評価

滋野 莉子¹ 山田 道洋² 山口 通智² 菊池 浩明¹ 坂本 真樹³

概要: CAPTCHA は, コンピュータと人間を区別するための完全自動化された公開チューリングテストであり, コンピュータエージェントによる不正行為を防止する重要な役割を果たしている. しかし, コンピュータの AI 技術が発展した現在, 既存の CAPTCHA が破られていることが大きな課題となっている. 機械判別が不可能な新しい CAPTCHA の生成が求められている. そこで, 本研究では, 擬態語や擬音語であるオノマトペを応用した「オノマトペ CAPTCHA」を提案する. オノマトペ CAPTCHA の精度と課題を明らかにし, ユーザビリティの向上を目的とする. 既存の代表的な CAPTCHA であるパズル型, 画像処理型と性能を比較する.

Evaluation and Development of Onomatopoeia CAPTCHA

RIKO SHIGENO¹ MICHIMIRO YAMADA² MICHITOMO YAMAGUCHI² HIROAKI KIKUCHI¹
MAKI SAKAMOTO³

1. はじめに

CAPTCHA(Completely Automated Public Turing Test To Tell Computers and Humans Apart) は, コンピュータと人間を区別するための完全自動化された公開チューリングテスト [1] である. コンピュータエージェントによる不正が深刻な課題となっている中, セキュリティー上で重要な役割を果たしている. CAPTCHA の必要条件は,

- (1) 人間が容易に解ける.
- (2) コンピュータには解くことができない.
- (3) 問題を自動生成できる.

ことである. しかし, コンピュータによる, 深層学習が発展した現在, 既存の CAPTCHA の多くが条件 (2) を満たさなくなっているという問題点がある. そこで, 機械判別が困難な新しい CAPTCHA の生成が求められている. 藤田ら

は, 3DCG 画像を用いた非現実画像 CAPTCHA[1] を提案している. 山口は機械合成文の不自然度相対識別問題に基づく CAPTCHA[2] を提案している.

本研究では, 擬態語や擬音語であるオノマトペ (Onomatopoeias) を応用した新たな「オノマトペ CAPTCHA」を提案する. オノマトペは普段人間が無意識に使用している概念であり, 音や響きから意味を推測できるので, 人間にとって解くのが簡単である. 一方でその解釈の原理はよく知られておらず, 機械的に解くのは困難と予測される.

本稿では, 提案方式が CAPTCHA の必要条件を満たすか明らかにするために, 出題方式を 4 形式作り, また既存のオノマトペと合成オノマトペを使用した問題の人間による正解率などを計測する. 幅広い被験者で評価するため, クラウドソーシングサイトを用いて被験者を集めて実験を行った. これにより, 人間にとって容易なオノマトペ CAPTCHA の形式を明らかにするとともに, 既存の CAPTCHA との性能比較も行う. またオノマトペ CAPTCHA の人間受入率と CAPTCHA が機械を受け入れる確率を明らかにし, 精度を明らかにする.

2. 基本定義

2.1 CAPTCHA について

「パズル型 CAPTCHA(図 1)」や「画像認証型

¹ 明治大学総合数理学部

School of Interdisciplinary Mathematical Science, Meiji University

² 明治大学大学院先端数理科学研究科

Graduate School of Advanced Mathematical Sciences, Meiji University,

³ 電気通信大学大学院情報理工学研究科

Graduate School of Informatics and Engineering, The University of Electro-Communications

true success

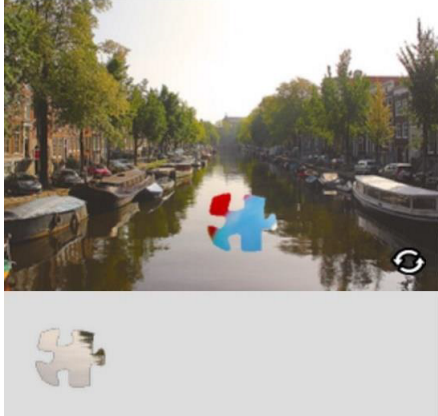


図 1 パズル型 CAPTCHA[6]



図 2 画像認証型 CAPTCHA[7]

CAPTCHA(図 2)」など様々な CAPTCHA が使用されている。パズル型 CAPTCHA は、図 1 の様にパズルの 1 ピースを絵や写真に適切にはめこむテストである。画像認証型 CAPTCHA は、与えられた記述に該当する写真を全て選択するテストである。図 2 では、「ワッフルの画像を選択してください」と出題されており、9 枚の写真の中から、ワッフルの写真の全て (1 行 3 列, 2 行 2 列, 2 行 3 列) の 3 枚を選べば正解である。

CAPTCHA が人間を受け入れる確率を HAR(Human Acceptance Rate) と定義する。一方, CAPTCHA が機械を受け入れる確率を MAR(Machine Acceptance Rate) と定義する. MAR が高いことは (1) 人に容易に解けることを, HAR が低いことは, (2) 機械には解けないことを表しており, これら MAR と 1-HAR はトレードオフの関係にある. HAR が高く, MAR が低いとき精度が高い CAPTCHA といえる. HAR は本実験の正解率と, 実験結果から算出した理論値と定義する. MAR は攻撃者の持つ能力に依存し, ブルートフォース攻撃と google 翻訳攻撃とでは確率が異なる。

2.2 オノマトペについて

オノマトペは、ものの音や声をまねた擬態語 (ざわざわ,

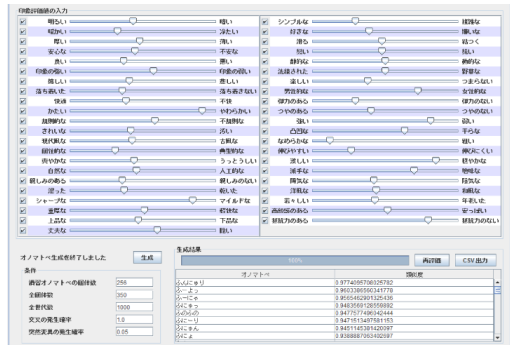


図 3 オノマトペ合成の実行例 [3]

じょきじょきなど), あるいは状態をまねた擬音語 (てきばき, きらきらなど), 擬声語 (わんわん, おぎゃー) である。一般語彙と比べると臨場感に溢れ、繊細な表現を可能としているという特徴がある。また、日本語は他の言語と比べてオノマトペの種類が圧倒的に多いと言われている [3]。小松らは、オノマトペの持つイメージを数値化し、オノマトペの持つイメージを客観的に表現するようなシステムを生成する研究をしている [4]。清水らは、ユーザーの入力した印象評価値に適合した音韻と形態を持つオノマトペ表現から、ユーザーの感性的印象に適合したオノマトペを生成するシステムを開発している [3]。図 3 にオノマトペを合成した例を示す。図 3 では「やわらかい」、「弱い」、「なめらかな」、「抵抗力のない」のパラメータを大きく設定し「ふんにゆり」、「ふーよっ」、「ふーにゃ」、「ふにゅっ」などのオノマトペが生成されている。このようにオノマトペにはさまざまな先行研究があるが、未だにオノマトペの明確な規則性は明らかにはされていない。

3. 提案方式

オノマトペ CAPTCHA はオノマトペの意味を正しく判断する問題を使用した CAPTCHA である。普段人間は正確な語でなくても、音のリズムや響きからそのオノマトペの意味を正しく理解することができる。しかしオノマトペには明確な規則性がないために、機械に意味を解釈するのは困難である。故にオノマトペは人間に容易 (条件 1) で、機械には困難 (条件 2) という点で CAPTCHA に適している。

3.1 オノマトペ辞書

オノマトペ CAPTCHA の問題生成するにあたって、まずオノマトペ辞書を作成した。オノマトペ辞書には、日本語オノマトペ辞典 [5] の意味分類の中から 20 項目を選出し、全 977 語を登録した。表 4 にその一部を示す。オノマトペ辞書の意味に対するオノマトペ数は一定ではない。その統計量を表 3 に示す。

本研究では、人間や機械について、オノマトペ CAPTCHA の解釈容易さや解答時間にどれだけ差が生じるかを調べる

表 1 オノマトベの解釈容易さの調査結果

	意味	総数 (人間予測)		総数 (Google 翻訳)	
		語数	%	語数	%
0	既知	359	37	262	27
1	未知	280	29	168	17
2	誤解	338	34	547	56
	計	977	100	977	100

表 2 オノマトベ辞書

タイプ	人間の予測	Google 翻訳	語数
A	0	1,2	210
B	1,2	0	112
C	1	1	59
その他 1	0	0	150
その他 2	2	2	190
その他 3	1	2	195
その他 4	2	1	61

表 3 実験に用いた辞書の意味に対するオノマトベ数

	語数
平均	48.85
最大	74
最小	11
総数	977

表 4 オノマトベ辞書の一部

意味 S	A	B	C
燃える	めらめら	ぼちぼち	ぼっ
走る	てくてく	どたん	さっき
笑う	うはうは	ころころ	うふっ
元気がない	がびーん	とぼとぼ	くしゅん
はやい	びゅん	ずんずん	さっき
ゆれる	ぶらぶら	へらへら	ゆきりゆきり
やわらかい	ぶよぶよ	なよなよ	くにゅくにゅ

ために、全オノマトベを人間と google 翻訳で評価した。その結果を表 1 に示す。ここで google でローマ字で表示された語は未知 (1) に分類した。また人間は著者のひとり (滋野) が判断した。

調査結果に基づいて、オノマトベを表 2 で定義した A, B, C の 3 タイプに分類した。A, B, C それぞれの例を表 4 に示す。例えば、A に分類した「めらめら」は、人間は正しい意味「燃える」と理解したが、Google 翻訳では異なった意味「Glittering: 輝く」が表示された。

A は人間には理解でき、機械には解けないという CAPTCHA にとって最良な語である。逆に、B は、人間には理解できず、機械には解けるという CAPTCHA にとって最悪な語である。

3.2 オノマトベ合成

2.2 節で述べた、ユーザの感性的印象に適合したオノマトベを合成するシステム [3] を用い、合成したオノマトベを問題に使用した。本オノマトベ生成システムは、与えた複数

表 5 合成オノマトベの例

印象語	合成オノマトベ			
かたい	ごげりごげり	ぞんげっ	ごんげっ	ごきーっ
きれいな	さんあり	しゃんあり	ざっあり	ひんあ
滑る	きゅう	しゅり	しゅえ	きゅら
湿った	ねべ	にゅちよ	ぬっちよ	ねびや
脆い	ひうひう	きろ	しゃろり	しわり
柔らかい	ふんにゅり	ふーにゃ	ふにゅっ	ふにーり

の印象評価値に適合した音韻と形態を持つオノマトベを自動合成する。表 5 に、本システムに与えた印象語とその出力された合成オノマトベの例を挙げる。これらは、全て google 翻訳では解釈できない語であった (タイプ C, または、その他)。

3.3 問題形式

本研究では、問題形式による人間の受入率と、機械受入率の変化を調査し、最適なオノマトベ CAPTCHA の形式を明らかにするため 4 種類の問題形式のオノマトベ CAPTCHA を検討する。それぞれの問題形式とその例を以下に示す。

形式 1 (4 択問題)

意味を表す最適なオノマトベをひとつ選択する。

(例) 「歩く」にあてはまるオノマトベは?

- ・ばくばく
- ・ぼそぼそ
- ・とことこ
- ・ぎやはは (答え) とことこ

形式 2 (オノマトベ選択)

オノマトベから類似のオノマトベを選択する。

(例) 「ぐずぐず」と同じ意味のオノマトベは?

- ・ほくほく
- ・へらへら
- ・ほろほろ
- ・たらたら (答え) たらたら

形式 3 (複数選択)

問題の意味にあてはまるオノマトベを全て選択する。

(例) 「喜ぶ」にあてはまるオノマトベを全て選べ

- ・ばくばく
- ・うきうき
- ・ごによごによ
- ・らんらん (答え) うきうき, らんらん

形式 4 (画像選択)

辞書の意味をもとに画像検索をして画像に合うオノマトベを選択する

(例) 画像の意味と同じ意味のオノマトベを選べ

- ・べらべら
- ・きよろり
- ・どんちゃん
- ・るんるん (答え) るんるん



3.4 出題方法

問題の出題方式を $A, B, C, RA, RB, RC, SK, SKAll$ の 8 種類用意した。 A は選択肢のオノマトベが全て A であり、 B, C も同様である。 RA は正解のオノマトベが A であり、他 3 つの選択肢は A, B, C ひとつずつである。 RB, RC も同様である。 SK は合成オノマトベが正解であり、他 3 つの選択肢は A, B, C ひとつずつである。 $SKAll$ は、全ての選択肢が合成オノマトベである。問題形式 3 では、与える語は全て A で正解の数を 1, 2, 3, 4 それぞれ 3 問ずつ用意した。問題形式 4 のタイプ A では、問題画像の数を 1 枚から 3 枚をそれぞれ 3 問ずつ用意した。各問題の出題数を表 6 に示す。

オノマトベ CAPTCHA の問題を自動生成するアルゴリズムを以下に示す。

出題アルゴリズム 1 形式 1-出題方式 A

- (1) 意味 S と語 $x \in D_{s,a}$ をランダムに選ぶ。
- (2) $S \neq S1, S2, S3$ となる意味 $S1, S2, S3$ と選択肢 $y \in D_{s1,A}, Z \in D_{s2,A}, w = D_{s3,A}$ をランダムに選ぶ。
- (3) 意味 S とランダムな順で x, y, z, w を出題する。
 $B, C, SKAll$ の出題方式も同様である。

出題アルゴリズム 2 形式 1-出題方式 RA

- (1) 意味 S と語 $x \in D_{s,a}$ をランダムに選ぶ。
- (2) $S \neq S1, S2, S3$ 選択肢 $y \in D_{s1,A}, Z \in D_{s2,B}, w = D_{s3,C}$ からランダムに選ぶ。
- (3) 意味 S とランダムな順で x, y, z, w を出題する。
 RB, RC, SK の出題方式も同様である。

出題アルゴリズム 3 形式 2

- (1) 意味 S と $x_0, x \in D_{s,A}$ をランダムに選び、 x_0 を問題 x を正答にする。
- (2) (2) はアルゴリズム 1 と同じ。
- (3) 出題語 x_0 と選択肢 x, y, z, w をランダムに出題する。

出題アルゴリズム 4 形式 3

- (1) 意味 S とその補集合 \bar{S} について、 $1/2$ の確率で、 $x \in D_{s,a}$ と $x \in D_{\bar{s},a}$ どちらか選ぶ。
- (2) y, z, w に対しても (1) 同様とする。

出題アルゴリズム 5 形式 4

- (1) 意味 S で検索した画像上位 10 枚からランダムに 1 枚選ぶ。
- (2) $x \in D_{s,a}$ をランダムに選ぶ。
- (3) $S \neq S1, S2, S3$ 選択肢 $y \in D_{s1,a}, Z \in D_{s2,a},$

表 6 各問題の出題数

	形式 1	形式 2	形式 3	形式 4
A	3	3	12	9
B	3	3	0	3
C	3	3	0	3
RA	3	3	0	3
RB	3	3	0	3
RC	3	3	0	3
SK	3	3	0	3
$SKAll$	3	0	0	0
計	24	21	12	27

$w = D_{s3,a}$ からランダムに選ぶ。

(4) 意味 S とランダムな順で x, y, z, w を出題する。

4. 評価実験

4.1 実験目的

本実験目的は次の通りである。

実験 1 出題形式による機械受入率 MAR と人間受入率 HAR の差を明らかにする

実験 2 既存のオノマトベと合成語や認知度による精度の差

実験 3 既存の CAPTCHA と提案したオノマトベ CAPTCHA の精度の比較

実験 4 人間の属性による結果の変化

年代・性別・職業など多種多様な被験者に対して実験を行い、属性の違いによって誤差率に差が生じるかを調査する。さらに、問題形式による、人間の受入率の差を計測する。

4.2 実験方法

本研究では、クラウドソーシングサイトにて募集した 54 名と菊池研究室の学生 9 名、合計 63 名を被験者とした。被験者には、実験用サイトにて表 6 に示す形式についての計 84 問の問題に解答してもらい、その解答内容と解答時間を記録した。また、実験終了後に 4.3.4 節のアンケート調査を行った。

4.3 実験結果

4.3.1 実験 1 出題形式の差

問題形式 1, 2, 3, 4 による実験結果と MAR を表 7 に示す。ブルートフォース攻撃による機械受入率を $MAR_{b,google}$ 翻訳による機械受入率を MAR_g と定義する。実験結果からの人間受入率を平均 HAR、理論値による人間受入率を HAR* と表す。表 7 より、最も人間受入率が高かったのは、問題形式 4 の画像から意味を読み取り、オノマトベを答える形式であった。問題形式 1 も 0.8 を超える平均人間受入率で、問題形式 2 と問題形式 3 はどちらも 0.6 後半の人間受入率だった。平均回答時間は問題形式による差はほとんどみられなかった。

表 7 形式の違いによる実験結果

形式	1	2	3	4
N	24	21	12	27
MAR_b	1/4	1/4	1/16	1/4
MAR_g	15/32	5/14	1/16	1/4
平均 HAR	0.845	0.668	0.680	<u>0.852</u>
HAR*	0.979	-	0.704	-
標準偏差	0.092	0.108	0.163	0.076
平均解答時間 (秒)	00:07	00:08	00:07	00:07
最大解答時間 (秒)	03:47	01:33	02:44	05:22

表 8 形式 3 の解答数による HAR

正解数	1	2	3	4
N	3	3	3	3
HAR*	0.704	0.699	0.694	0.688
平均 HAR	0.540	0.709	0.720	<u>0.751</u>
平均解答時間 (秒)	00:07	00:07	00:07	00:08
最大解答時間 (秒)	00:51	00:59	00:51	02:44

MAR_b は, 表 2 のオノマトペ辞書で B に分類される単語が, google 翻訳で正しく意味が解釈される時に, CAPTCHA が機械を誤って受け入れると仮定して算出した. すなわち, 機械が正解する確率が google 翻訳による MAR である. 例えば, 形式 1 で選択肢がタイプ A, B, C, A で答えが A の場合, 4 択の中で B が答えでないことはわかるので正解する確率は $1/3$ となる. 形式 1 と形式 2 は全ての出題形式の HAR を算出し, その平均を表 7 に示している. 形式 3 では, 全てのオノマトペにタイプ A を使用しているため, ブルートフォースと google 翻訳の MAR は変わらない, すなわち, $MAR_b = MAR_g$ である. 形式 4 では問題が画像のため, オノマトペの意味を理解できても画像の意味を理解しないと正解することはできないため, google 翻訳はブルートフォースと同じ MAR となる.

問題形式 3 では, 複数選択の正解の選択肢の数を 1 つから 4 つまで 3 間ずつ用意した. 正解の選択肢の数ごとの平均 HAR と, 平均解答時間を表 8 に示す. また不正解の解答時に, 被験者が実際の正解の選択肢の数に対していくつの選択肢を選択したかを表 9 に示す. 表から正解の選択肢の数が 1 つの時の人間受入率が最も低い. また正解の選択肢の数が 1 つの時に, 選択肢を 2 つ選んでしまう被験者が極めて多かった.

表 10 は形式 4 の問題画像を変化させた際の平均人間受入率と平均解答時間である. 表から, 画像の枚数が少ないほど平均人間受入率が高くなり, 平均解答時間も短くなった.

4.3.2 実験 2 合成語と辞書

出題形式 A, B, C ごとの結果を表 11 に示す. 人間受入率は高い順に C, A, B となり, 解答時間は短い順に C, B, A となった. 人間には困難だと思われ, Google 翻訳で正しい意味で翻訳された B が圧倒的に HAR が低かった.

既存のオノマトペと合成オノマトペとの平均人間受入率

表 9 形式 3 間違い

解答個数 \ 正答個数	1	2	3	4
N	87	55	53	47
1	0	21	4	6
2	78	10	36	11
3	7	23	4	30
4	2	0	9	0

表 10 形式 4 画像の枚数ごとの正解率

画像の枚数	平均時間 (秒)	平均 HAR
1	5	0.937
2	10	0.884
3	11	0.783

表 11 A, B, C による人間受入率

	A	B	C
N	15	9	9
平均 HAR	0.831	0.757	<u>0.858</u>
平均解答時間 (秒)	00:08	00:07	00:06
最大解答時間 (秒)	02:56	01:12	00:58

表 12 オノマトペ辞書と合成語オノマトペの比較

形式	辞書	合成	
		SK	$SKALL$
平均 HAR			
1	<u>0.891</u>	0.757	0.656
2	<u>0.705</u>	0.444	-
4	<u>0.894</u>	0.513	-
時間 (秒)			
1	00:07	00:08	00:08
2	00:08	00:09	-
4	00:07	00:09	-

と平均解答時間をそれぞれ表 12 に示す. SK が選択肢に合成語オノマトペを 1 つ使用した問題で, $SKALL$ が選択肢の全てが合成語オノマトペの問題である. 合成オノマトペを使用した形式 1, 2, 4 では既存のオノマトペと比較し, どの形式も平均人間受入率は低く, 平均で 0.25 低い. また, 表 12 より選択肢の 1 つに合成オノマトペを使用した問題 (SK) は, 4 つの選択肢すべてに合成オノマトペを使用した問題 ($SKALL$) より, 0.11 低い. 以上より, 合成オノマトペは機械受入率を下げるが同時に, 人間受入率も下げることが示された.

4.3.3 実験 3 既存の CAPTCHA 方式の比較

既存の CAPTCHA と提案オノマトペ CAPTCHA との平均人間受入率と平均解答時間の比較結果を表 13 に示す. 既存の CAPTCHA であるパズル型 CAPTCHA と画像認証型 CAPTCHA は 2016 年 8 月に 200 名の被験者に対して評価した. 表 13 は, これら既存の CAPTCHA の実験結果と本研究の問題形式 1 の平均正解率と平均解答時間である.

実験結果から, オノマトペ CAPTCHA は他 2 つの CAPTCHA と比べて平均人間受入率は高く, 解答時間も短

表 13 既存の CAPTCHA との比較結果

	平均解答時間 (秒)	標準偏差	平均 HAR(%)	標準偏差
オノマトベ	7	9.92	<u>89.1</u>	0.092
パズル	9.44	5.23	88	0.346
画像	18.07	11.65	81	0.388

表 14 補正をした結果

	MAR _b	補正後		
		問題数 x	MAR _b	解答時間 (秒)
オノマトベ	1/4	4.5	0.595	31.5
パズル	1/90	1.38	<u>0.838</u>	13
画像	1/512	1	0.810	18.07

く見える。しかしコンピュータのブルートフォース攻撃による受入率が異なるため、公平な評価ではない。例えば、オノマトベ CAPTCHA は、 $MAR_b = 1/4 > 1/512$ であり、画像 CAPTCHA よりはるかに機械に対する安全性が低い。そこで、機械に対する安全性を揃える為に、各問題を x 回繰り返すことを考える。3 方式の MAR_b を等しくするには、

$$\left(\frac{1}{4}\right)^{x_o} = \left(\frac{1}{90}\right)^{x_p} = \frac{1}{512}$$

を解いて、

$$x_o = \log_4 512 = 4.5$$

$$x_p = \log_{90} 512 = 1.38$$

である。従って、等価な安全性を得るにはオノマトベ CAPTCHA を 4.5 回繰り返す必要があり、8 秒 \times 4.5 の平均解答時間と $0.891^{4.5} = 0.595$ の平均人間受入率に補正される。以上の補正結果を表 14 に示す。

それぞれの CAPTCHA のブルートフォース攻撃の確率の求め方は以下に示す。

(1) オノマトベ CAPTCHA のブルートフォース攻撃

表 14 のオノマトベ CAPTCHA は形式 1 の 4 択問題の場合である。4 つの選択肢から答えを 1 つ選択するため MAR_b は $1/4$ となる。

(2) パズル型 CAPTCHA のブルートフォース攻撃
パズル型 CAPTCHA はパズルのピースをはめこむ CAPTCHA である。マスが 30×27 あり、正解の位置前後 9 マスが受入れられる。よってパズル型 CAPTCHA の MAR_b は、 $1/90$ となる。

(3) 画像認証型 CAPTCHA のブルートフォース攻撃

画像認証型 CAPTCHA が 9 枚の画像から正しいものを全て選択するため、 $MAR_b = 2^{-9} = 1/512$ となる。

(4) ブルートフォース攻撃を補正

それぞれの CAPTCHA でブルートフォース攻撃による受入率が異なるため、等しく補正をした。補正は上記の計算方法で導いた。

4.3.4 実験 4 被験者の属性の比較

本被験者に対して、以下の 5 項目についてアンケート調査を行った。

(1) 性別

表 15 アンケート結果

	N	平均正解数	平均 HAR	平均解答時間
男	29	64.03	0.760	9:47
女	34	66.56	<u>0.792</u>	9:53
10 代	8	60.88	0.725	11:27
20 代	11	65.55	0.780	10:47
30 代	26	66.35	0.790	<u>9:37</u>
40 代	15	66.53	<u>0.792</u>	11:02
50 代	2	63.00	0.750	10:50
60 代	1	63.00	0.750	18:38
文系	43	65.60	<u>0.780</u>	10:45
理系	20	65.40	0.779	10:12
漫画 ($n \geq 10$)	22	64.27	0.765	10:44
漫画 ($n < 10$)	41	66.00	<u>0.786</u>	10:29
本 ($n \geq 10$)	18	65.17	0.776	10:21
本 ($n < 10$)	45	65.49	<u>0.780</u>	10:40

表 16 マッチと非マッチ

R \ Q	マッチ m	非マッチ w
マッチ m	p	$1 - q$
非マッチ w	$1 - p$	q

- (2) 年齢 (10 代から 60 代まで)
- (3) 文系か理系か
- (4) この 3 か月で読んだ漫画の数
- (5) この 3 か月で読んだ本の数

アンケート結果に基づく平均人間受入率と平均解答時間を表 15 に記す。表から、30 代 40 代が最も平均人間受入率が高く、平均解答時間も短かった。逆に 10 代の人間受入率が最も低かった。性別、文理、漫画と本の読んだ数の差による実験結果の違いはほとんどみられなかった。

4.4 機械による安全性

本研究では人間がオノマトベ CAPTCHA を解くための理論値を求めるため、形式 3 の複数選択問題を用いて、正しい判別を行う確率を求める。 Q と R をクエリーとレスポンスの確率変数とし、 m (マッチ) か w (誤り) のどちらかの値をとる。この時、意味の正しいオノマトベを正しいと判断する確率を

$$p = Pr(R = m | Q = m)$$

正しくないオノマトベを正しくないと判断する確率を

$$q = Pr(R = w | Q = w)$$

とする。

形式 3 は複数選択問題のため、各々のオノマトベに対して正しいか正しくないかを確認している。表 17 では形式 3 で正しいオノマトベを解答した数、表 18 では誤ったオノマトベを選択した数を示している。この総数に対する割合が表 17 では p 、表 18 では q となる。

表 17 複数選択正しい選択

形式 3 全ての問題					
選択肢番号	1	2	3	4	計
N	378	567	441	504	1890
選択した数	346	520	405	468	1739
形式 3 で選択肢が 1 つのもののを除いた問題					
N	378	504	378	441	1701
選択した数	346	457	342	405	1550

表 18 複数選択間違った選択

形式 3 全ての問題					
選択肢番号	1	2	3	4	計
N	378	189	315	252	1134
選択した数	254	177	307	252	990
形式 3 で選択肢が 1 つのもののを除いた問題					
N	189	63	189	126	567
選択した数	143	63	189	126	521

本実験では、解答数が 1 つの問題は極めて人間受入率が低かった。正しい答えを 1 つ選択してさらに誤った答えを選択し、2 つ選択する被験者が極めて多かったことから、複数問題と表示されているため答えが一つではないという心理的影響を受けたと考えられる。よって形式 3 の答えが 1 つの問題を抜いた数値を使用する。表より

$$p = 1550/1701 = 0.911$$

$$q = 521/567 = 0.918$$

を得る。

4 択問題では 4 つの選択肢から 1 つを選択するため、形式 1 の人間受入率 MAR の理論値は、

$$\begin{aligned} Pr(\text{形式 1 を正解する}) &= 1 - (1 - p) \cdot (1 - q^3) \\ &= 1 - (1 - 0.911) \cdot (1 - 0.918^3) \\ &= 0.979 \end{aligned}$$

と予測できる。

理論値 HAR* は 0.979 となったが、本実験の形式 1 の人間受入率は 0.845 となった。理論値 HAR* と比較して人間受入率は 0.134 低い。形式 3 ではタイプ A のオノマトペのみを使用している。一方、形式 1 では A, B, C の他に合成オノマトペを混入している。表 12 より合成オノマトペを抜いた形式 1 の人間受入率は 0.891 であり、人間に理解しやすいオノマトペが多かったため、この差が生まれたと考えられる。

4.5 解答時間のヒストグラム

ユーザごとの HAR と総解答時間を図 4 に示す。総解答時間が最も長いユーザでも HAR が 0.75 程度であることなどから、総解答時間と HAR の間に相関はないと考えられる。

年代別の総解答時間の分布と HAR の分布それぞれ図 5、図 6 に示す。どちらも分布に年代による大きな差は示され

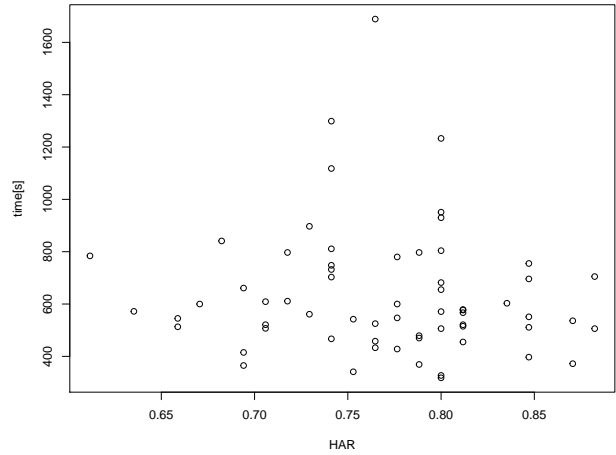


図 4 ユーザごとの HAR と総解答時間

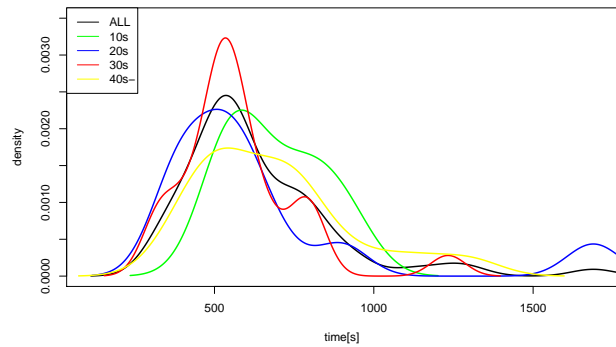


図 5 年代ごとの総解答時間の分布

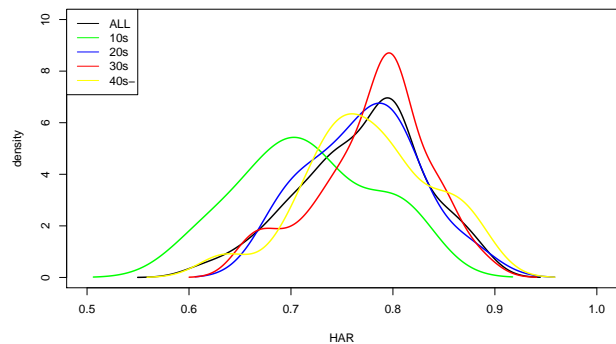


図 6 年代ごとの HAR の分布

なかった。

4.6 問題・出題形式の差の確率検定

出題・問題形式の違いによる解答時間・HAR への影響を確かめるために t 検定を行った結果を表 19 表 20 にそれぞれ示す。形式 1, A をそれぞれ基準、 $p < 0.05$ を有意とした場合、全ての問題・出題形式について HAR については形

表 19 問題形式についての検定

問題形式	平均	statistic	p-value
解答時間についての検定			
(intercept)	00:07	-	-
2	00:08	-2.169	0.030*
3	00:07	-0.717	0.472
4	00:07	-0.767	0.442
HAR についての検定			
(intercept)	0.845	-	-
2	0.668	9.734	6.90e-17*
3	0.680	6.904	5.05e-10*
4	0.852	-0.477	6.33e-01*

表 20 出題形式についての検定

問題形式	平均	statistic	p-value
解答時間についての検定			
(intercept)	00:08	-	-
B	00:07	2.637	0.008*
C	00:06	2.977	0.002*
RA	00:08	1.141	0.253
RB	00:07	1.721	0.085
RC	00:06	3.337	0.000*
SK	00:09	0.028	0.977
HAR についての検定			
(intercept)	0.823	-	-
B	0.759	2.637	1.89e-04*
C	0.856	2.977	1.48e-01*
RA	0.885	1.141	1.08e-03*
RB	0.872	1.721	1.07e-02*
RC	0.864	3.337	9.80e-03*
SK	0.597	0.028	4.72e-20*

式 1,A に対してそれぞれ有意差が見られ, 問題・出題形式によって難易度が異なっていたということが示された。

4.7 考察

本実験より, 人間にとってオノマトペは年齢や性別など属性に関係ないことが明らかになった。また問題形式 1~4 の結果から, イメージを想像しやすい画像からオノマトペを選ぶ問題形式 4 の HAR が最も高かったのは, 人間はオノマトペを明確な意味ではなく, イメージで把握しているからだと考えられる。

また, 問題形式 3 で解答個数が 1 つの時に, 人間受入率が低く, 2 個以上選択してしまう被験者が多かったのは, 「すべて選べ」という設問文による心理的影響が大きいと考えられる。オノマトペ辞書や合成オノマトペの結果を比較したところ, タイプ B や合成オノマトペの人間受入率が低かったことから, 普段馴染みのないオノマトペは人間にはイメージしにくく少し困難である。

既存の CAPTCHA に比べてオノマトペ CAPTCHA は人間受入率が高く, 解答時間も短かった。しかし, ブルートフォースの耐性を考慮すると, 提案オノマトペ CAPTCHA

にはまだ実用的な問題が残っている。コンピュータに受け入れられないようにすることをオノマトペ CAPTCHA の今後の課題とする。

5. おわりに

本研究では「オノマトペ CAPTCHA」を提案し, 人間受入率と機械受入率の性能評価を行った。実験結果より, オノマトペ CAPTCHA は人間に優しくストレスなく解くという観点では, 既存の CAPTCHA よりも優れているといえる。しかし機械に対する安全性を考えるとまだ十分な CAPTCHA とは言えない。今後は, コンピュータに対する受入率を下げ, 視覚障害の方や日本語以外にもオノマトペ CAPTCHA ができるようにすることも目標である。

謝辞

本研究を行うにあたり, さまざまな方にご協力いただきました。電気通信大学大学院情報理工学研究科の紺野倭人氏, 鈴木航平氏には, 実験に使用した合成オノマトペのデータを提供していただき, オノマトペについてのご助言をいただきました。また, 明治大学総合数理学部先端メディアサイエンス学科の中村聡史准教授, 小松孝徳准教授には, オノマトペのご助言をしていただきました。最後に, 実験に協力してくださった被験者に感謝します。

参考文献

- [1] 藤田真治, 池谷勇樹, 西垣正勝: 非現実画像 CAPTCHA: 常識からの逸脱を利用した 3DCG 画像 CAPTCHA, 情報処理学会論文誌, 56(12), pp.2324-2336, 2015.
- [2] 山口通智, 岡本健, 菊池浩明: 機械合成文の不自然度相対識別に基づく CAPTCHA の提案, 情報処理学会論文誌, 56(9), pp.1834-1835, 2015.
- [3] 清水祐一郎, 土斐崎龍一, 鍵谷龍樹, 坂本真樹: ユーザの感性的印象に適合したオノマトペを生成するシステム, 人工知能学会論文誌, 30(1), pp.319-330, 2015.
- [4] 小松孝徳, 秋山広美: ユーザの直観的表現を支援するオノマトペ意図理解システム, 電子情報通信学会論文誌, vol.J92-A(11), pp.752-763, 2009.
- [5] 小野正弘: 擬音語・擬態語 4500 日本語オノマトペ辞典. 2007.
- [6] Capy パズルキャプチャ (https://www.capy.me/jp/products/puzzle_captcha).
- [7] ReCAPTCHA demo (<https://www.google.com/recaptcha/api2/demo>)