明治大学総合数理学部

2017 年度

卒　業　研　究

環境雑音を用いた
音声 CAPTCHA の認識実験

学位請求者　先端メディアサイエンス学科

二谷太郎

目次

1　はじめに 1

1.1　研究背景 1

1.2　CAPTCHA 1

1.3　既存の CAPTCHA の問題点 3

1.4　研究目的 3

2　環境雑音を用いた音声 CAPTCHA 3

2.1　提案方式 3

2.2　実験 1　人間受理率 4

2.2.1 実験目的 4

2.2.2 実験方法 4

2.2.3 実験結果 5

2.3　実験 2　機械受理率 6

2.3.1 実験目的 6

2.3.2 実験方法 7

2.3.3 実験結果 7

2.4　考察 10

3　音声 CAPTCHA に有用なボイスパターンの調査実験 10

3.1　調査実験について 10

3.2　実験内容 10

3.3　実験結果 11

3.4　考察 11

4　おわりに 15

謝辞 15

参考文献 16

付録 A　近赤外戦分光法 NIRS を用いたストレスの強さの定量化

A.1 はじめに 17
A.1.1 研究背景 17
A.1.2 研究目的 17
A.1.3 近赤外線分光法 NIRS 17
A.2 実験 18
A.2.1 実験概要 18
A.2.2 データ方式 19
A.2.3 ストレス指標の提案 20
A.2.4 実験結果 21
A.2.5 考察 22
A.3 おわりに 22

1. はじめに

1.1. 研究背景

インターネットの普及により，様々なサービスをオンライン上で受けることができるようになった．ウェブ上でのショッピングを始めとして，人間と人間がウェブを介して取引をするようになっている．しかし，取引をしている上で取引相手が本当に人間であるかどうか不明である点が懸念されており，悪意を持った利用者がボットにサービスを不正利用させるという事件が多数挙げられている．

1.2. CAPTCHA

ボットによるサービスの不正利用を防ぐために，CAPTCHA(Completely Automated Public TuringTest To Tell Computers and Humans Apart)が使用されるようになった．図 1 に CAPTCHA の構成図を示す．CAPTCHA とは,サービス利用者に対して簡単な問題を投げかけることによって人間かロボットかを判別する完全に自動化された公開チューリングテストである．CAPTCHA の必要条件は以下の通りである．

1. 人間にとって解くのが容易である
2. コンピュータにとって解くのが困難である
3. 問題の自動生成が可能である

CAPTCHA には歪んだ文字画像を表示してその文字を判別させる文字画像判別型の CAPTCHA，非現実の 3DCG を用いた画像 CAPTCHA やオノマトペを用いた CAPTCHA などを始めとした視覚型のものと[1][2][3]，聴覚に訴えて雑音を背景にした状況下のアルファベットを判別させるような聴覚型の CAPTCHA がある[4][5][6]．図 2 に主流である文字画像判別型 CAPTCHA の例[7]を，図 3 に画像選択型の CAPTCHA の例[8]を示す．図 3 の画像選択型の CAPTCHA では，指定する対象の画像を 1 つ以上選択させることで判別を行う．

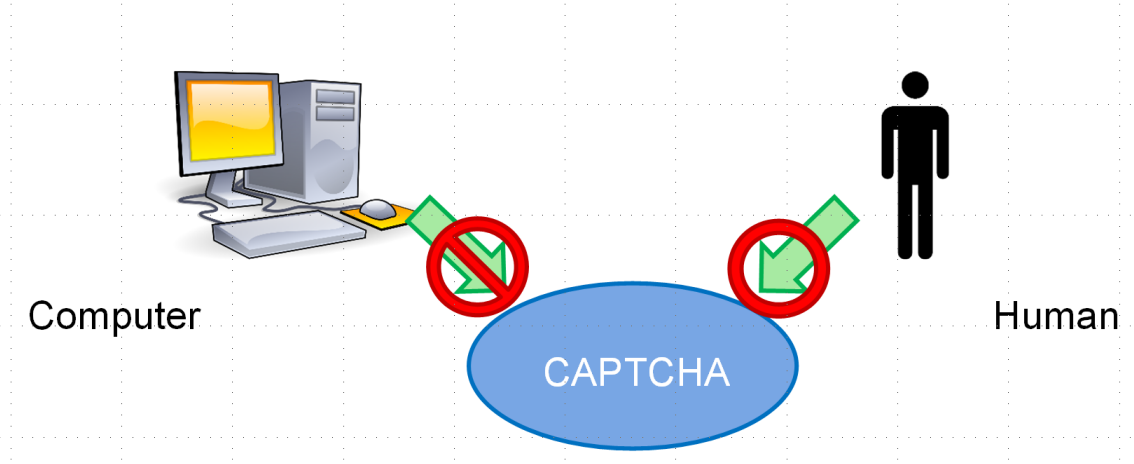

図 1　CAPTCHA の構成図

図 2　文字画像判別型 CAPTCHA の例

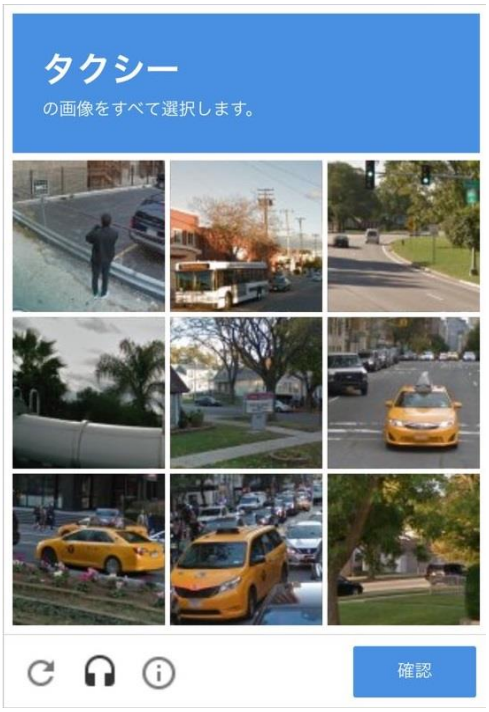

図 3　画像選択型 CAPTCHA の例

1.3. 既存の CAPTCHA の問題点

現在の CAPTCHA には様々な問題点が挙げられる．文字画像判別型の CAPTCHA の問題点は，安全性である．近年のボットの技術に対応できなくなっており，OCR を導入したボットに突破されるケースが多くなっている．それに伴い，問題が複雑化して人間が解けないような問題を出題される場合も多く，CAPTCHA が満たされなくなってしまっている．

さらに，聴覚型の CAPTCHA は視覚に障害を持っている人が行うものが多いため，ただでさえ人間にとって容易である点が満たされづらい．その中で音声認識の発展につれて難聴化を繰りかしているため，より人間にとって解くのが困難になってしまっている．

1.4. 研究目的

本研究では，CAPTCHA の必要条件に基づいた新しい聴覚型の CAPTCHA の提案を行う．人間が様々な環境雑音（カフェの作業音や駅前の雑踏など）の中で会話することに慣れていることに着目し，今までの白色雑音やヒス雑音を用いたものとは異なり，環境雑音を用いた音声 CAPTCHA を提案する．人間にとっては，普段の環境下とあまり変わらないため解くことが容易であり，機械にとっては雑音の中に紛れている音や人々の会話と問題として出すテキストの判別がつかないことを期待する．

2. 環境雑音を用いた音声 CAPTCHA

2.1. 提案方式

提案方式を以下に示す．

1. カフェの作業音や駅前の雑踏の音を動画投稿サイトから取得し，2.5 秒程度の長さに切り取る．
2. 日本語の単語（2 文字～8 文字）を Open JTalk を使用して音声合成する．使う音声の種類は女性ボイス（平均よりやや高めのもの）を使用した．
3. 環境雑音と単語の音声を重ね，再生し，認識テストにより，人間であることを認識する．

図 4 に実装した実験ページの実行例を示す．

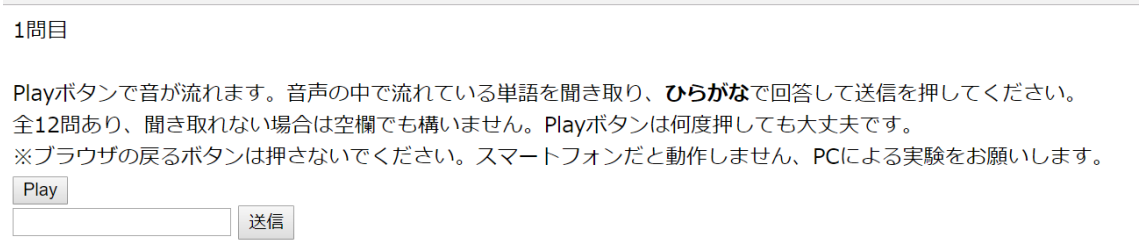

図 4　実験ページの実行例

2.2.　実験 1　人間受入率

2.2.1.　　実験目的

CAPTCHA が人間を正しく受け入れる確率 HAR(Human Acceptance Rate)を求め，提案方式の精度を明らかにする．

2.2.2.　　実験方法

2.1 節にて提案した CAPTCHA を単語の文字数が 2, 4, 6, 8 文字のものを各 3 問ずつ(聞きなじみのある単語やあまり聞きなじみのない単語を含む)の全 12 問の実験を行った．表 1 に使った単語のリストを示す．明治大学総合数理学部の学部生の合計 44 名を被験者として 11 月にウェブ上で実験を行った．本実験では，被験者が正しく回答した数を *sucH*，回答した問題数を *queH* とし，以下のように HAR を定義する．

$$\mathrm{HAR} = sucH \ / \ queH$$

表 1　単語リスト

2 文字	あさ	ひじ	もず
4 文字	おはよう	らーめん	かんかつ
6 文字	いちごいちえ	すいへいせん	だいだんえん
8 文字	かんこんそうさい	とくがわつなよし	あすぱらぎんさん

2.2.3.　実験結果

男女別，年齢別の HAR を表 2 に，文字数別の HAR を表 3 に示す．

表 2 より，19 歳の平均 HAR は 0.750，23 歳の平均 HAR は 0.682 となり，年齢の上昇とともに HAR の低下が見受けられた．年齢のデータが足りないのでこの実験から年齢と正答率の関係を判断するのは難しいが，聴力と HAR が関係する可能性が考えられる．

表 3 より，回答文字数 8 文字の HAR が 0.810 で最も高かった．文字数が少ない場合，前後の語から補完する力があまり働かないため，文字数すら誤答するケースが多々あった（例:ひじ→にんじん）．しかし，8 文字程度のテキストになると補完能力が強く働き，未知の単語でもある程度は正答できると考えられる．

表 2　男女別，年齢別の HAR

		N	平均	標準偏差	最大 HAR	最小 HAR
性別	男	35	0.705	0.140	0.917	0.333
	女	9	0.732	0.094	0.917	0.583
年齢	19	1	0.750	0.000	0.750	0.750
	20	8	0.729	0.176	0.917	0.333
	21	10	0.725	0.112	0.917	0.500
	22	14	0.708	0.069	0.833	0.583
	23	11	0.682	0.170	0.917	0.333
計		44	0.710	0.132	0.917	0.333

表 3　文字数別の HAR

文字数	平均 HAR
2	0.772
4	0.659
6	0.598
8	0.810

2.3. 実験 2　機械受入率

2.3.1. 実験目的

CAPTCHA が機械を誤って受け入れる確率 MAR(Machine Acceptance Rate)を求め，提案方式の精度を明らかにする.

2.3.2. 実験方法

2.2 節で使用した 12 問の CAPTCHA と単語の音声データのみの 12 問の計 24 問を Siri と Google の音声認識音声入力にそれぞれ 10 回ずつ行った．本実験では，機械が正しく認識した数を *sucM*, 認識した問題数を *queM* とし，以下のように MAR を定義する．

$$\mathrm{MAR} = sucM \ / \ queM$$

2.3.3. 実験結果

認識システム毎の MAR を表 4 に示す.

表 4　認識システム毎の MAR

認識システム	単語単体（MAR）	CAPTCHA（MAR）
Siri	0.450	0.000
Google	0.575	0.058

表 4 より，CAPTCHA の MAR は十分に低い．Siri の場合は全ての CAPTCHA でテキストを認識することすらなかった．Google の場合は CAPTCHA のテキストを認識することはあったが多くは誤認した．Google の単語単体に対する音声入力の認識結果の一覧を表 5 に，Google の CAPTCHA に対する音声入力の認識結果の一覧を表 6 に示す．かっこ内に認識した回数を示す.

表 5 と表 6 より，単語単体で正確な認識をしているものに対して CAPTCHA にしたものは全て正答率が下がっていることが分かった．テキストが短いものは同じ誤認をすることが多く，テキスト数が長くなればなるほど，著しい誤認をすることが分かった．単語単体では 8 文字において全て正確に認識しているのに対して CAPTCHA にすると全て誤認することが見受けられるため，8 文字程度までテキストを長くすると，雑音を入れることにより一部のみ認識できた語から間違った補完をしたり，複数文節と勘違いしたりする確率が増え，機械に認識されにくいことが分かった.

表 5　Google の単語単体に対する音声入力の認識結果の一覧

あさ	**朝**（10）
ひじ	釣り（5），7（2），チビ（2），次（1）
もず	マジ（6），なぜ(4)
おはよう	**おはよう**（10）
らーめん	**らーめん**（3），大画面（3），誰（2），画面（2）
かんかつ	**管轄**（2），観察（7），暗殺（1）
いちごいちえ	**一期一会**（10）
すいへいせん	産総研（4），戦争ゲーム（2），電車遅延（1），先生（1），戦争（1），戦争犬（1）
だいだんえん	**大団円**（4），ダイダロス（4），鯛ラーメン（2）
かんこんそうさい	**冠婚葬祭**（10）
とくがわつなよし	**徳川綱吉**（10）
あすぱらぎんさん	**アスパラギン酸**（10）

表 6　Google の CAPTCHA に対する音声入力の認識結果の一覧

あさ	**朝**（2），認識しない（8）
ひじ	次（2），ちび（1），認識しない（7）
もず	マック（1），認識しない(9)
おはよう	**おはよう**（2），若菜（3），お花（3），高山（1），わかる（1）
らーめん	**らーめん**（2），誰に（3），誰（2），画面（2），ランニング（1）
かんかつ	暗殺（9），アイカツ（1）
いちごいちえ	**一期一会**（1），一番強い（5），すごいする（2），素晴らしい（1），1 月（1）
すいへいせん	愛媛県（3），YouTube（3），南海ホークス（1），岐阜県（1），電源コード（1），犬ゲーム（1）
だいだんえん	画面（2），ダイナミック（2），アリナミン（2），ダイレンジャー（1），排卵日（1），海外人気（1），チャイナムーン（1）
かんこんそうさい	おそ松さん（3），iphone 4 サイズ（3），三陽商会（2），エクササイズ（1），岡山正社員（1）
とくがわつなよし	はやくなるし（1），鳥田市（1），松原市（1），伊勢志摩市（1），松村治樹（1），鹿児島です（1），大阪松原市（1），小松菜レシピ（1），山梨マルス（1），ヤマカガシマムシ（1）
あすぱらぎんさん	カラカラ進化（6），バドミントン（1），ここから近いスーパー（1），埼玉銀行（1），三原じゅん子（1）

2.4. 考察

2.2 節と 2.3 節の実験結果より，8 文字のテキストの問題は HAR が高く，MAR が低いことが明らかになった．この特性を利用することでより精度の高い CAPTCHA を作ることができると考えられる．

3. 音声 CAPTCHA に有用なボイスパターンの調査実験

3.1. 調査実験について

群衆雑音を用いた音声 CAPTCHA を開発するにあたり，使用するボイスの種類は何がよいかを把握したかったため 9 月にそれを探る実験を行った．

3.2. 実験内容

16 パターンのデモサウンドを用意し，2.2 節で行ったような実験を行った．16 パターンの内訳は以下の通りである．

- 音声の性別（男女）
- 音声の高低
- 音声テキストの長さ（4 文字以下と 4 文字以上）
- ノイズと音声のバランス（音声の音量の大小）

以上の各 2 通りずつの計 16 種類のデモサウンドを生成した．なお，使ったテキストは 16 パターン全て別のものを使用した．表 7 に使用したテキストとパターンを示す．パターン列の 3 ケタの数字の百の位は音声の高低（0 が低い），十の位はテキストの長さ（0 が 4 文字以下のテキスト），一の位は音量バランス（0 が音声が小さい）を表す．

基本的には 2.1 節と作問方法は同じだが，この実験では音声合成ソフトとしてフリーソフトである SofTalk を使用し，音声と雑音を合成するソフトとして同じくフリーソフトである Audacity を使用した．

明治大学総合数理学部の学部生と院生を始めとして，2017 年度の明治大学のオープンキャンパスに参加した高校生のうち菊池研究室に見学しに来た学生，親類や他大学の友人に 103 名を対象として実機で実験を行った．

表 7　使用したテキストとパターン

パターン	女（f）	男（m）
000	ぱそこん	たんぼ
001	とうだい	やきにく
010	ぜったいぜつめい	ぶたにしんじゅ
011	べんけいのなきどころ	やけいしにみず
100	とけい	おはよう
101	たまご	やきゅう
110	めいじだいがく	はなよりだんご
111	ほんのうじのへん	まんじょういっち

3.3. 実験結果

表 8 に実験結果（正答率（%））を，図 4，図 5，図 6，図 7 に要素ごとに抽出した正答率のグラフを示す.

全体を通しての正答率は 44.4%であり，最も正答率の高いものは f111（女性高音テキスト長音量大）であり，最も正答率の低いものは m000（男性低音テキスト短音量小）であった.

3.4. 考察

聞き取りやすくなる傾向は，以下の通りであると考えられる.

- 音声の種類：女性の声
- 音声の高低：高音の声
- テキストの長さ：長い
- 音量のバランス：音声が大きい

実際に実験を行った際に感想を聞いたところ，この傾向に沿っているものほど解きやすいと感じたという意見が多かった．今後作問する際にはこの傾向に沿った音声を使用すれば人間に対する受け入れ率が上がると考える.

表 8　夏課題の実験結果（正答率（%））

ぱそこん	f000	46.6
とうだい	f001	26.2
ぜったいぜつめい	f010	60.1
べんけいのなきどころ	f011	89.3
とけい	f100	22.3
たまご	f101	95.1
めいじだいがく	f110	73.8
ほんのうじのへん	f111	96.1
たんぼ	m000	0.0
やきにく	m001	60.2
ぶたにしんじゅ	m010	9.7
やけいしにみず	m011	31.1
おはよう	m100	21.4
やきゅう	m101	15.5
はなよりだんご	m110	5.8
まんじょういっち	m111	58.3

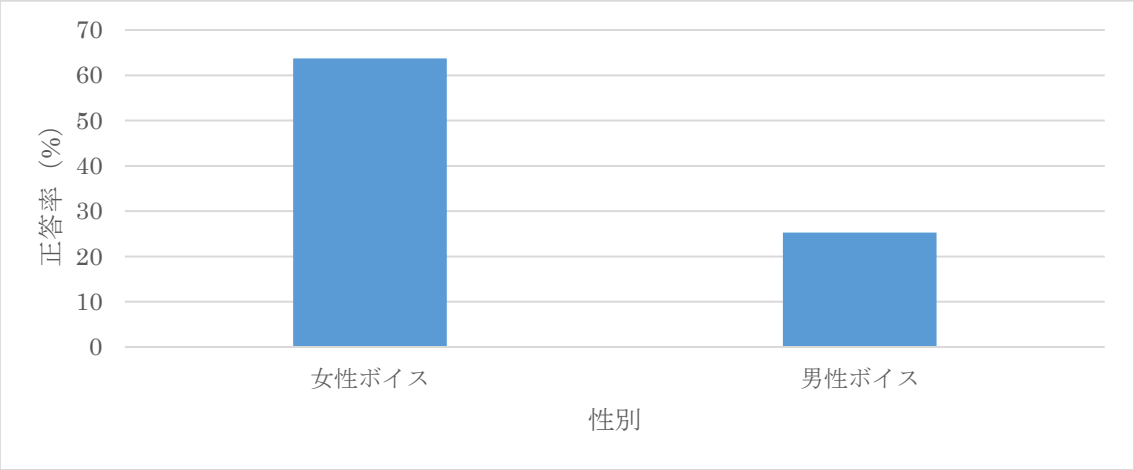

図 4　男女別正答率

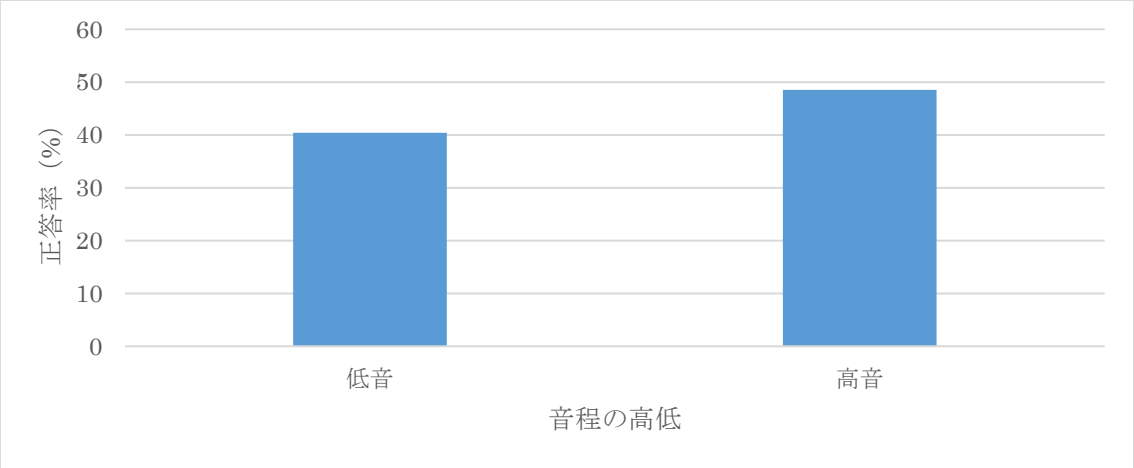

図 5　音声の高低別正答率

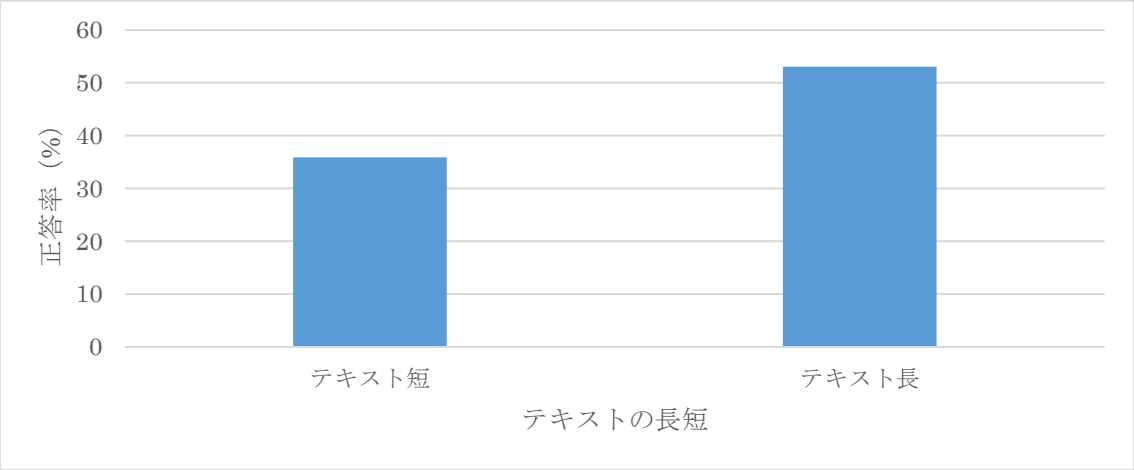

図 6　テキスト長短別正答率

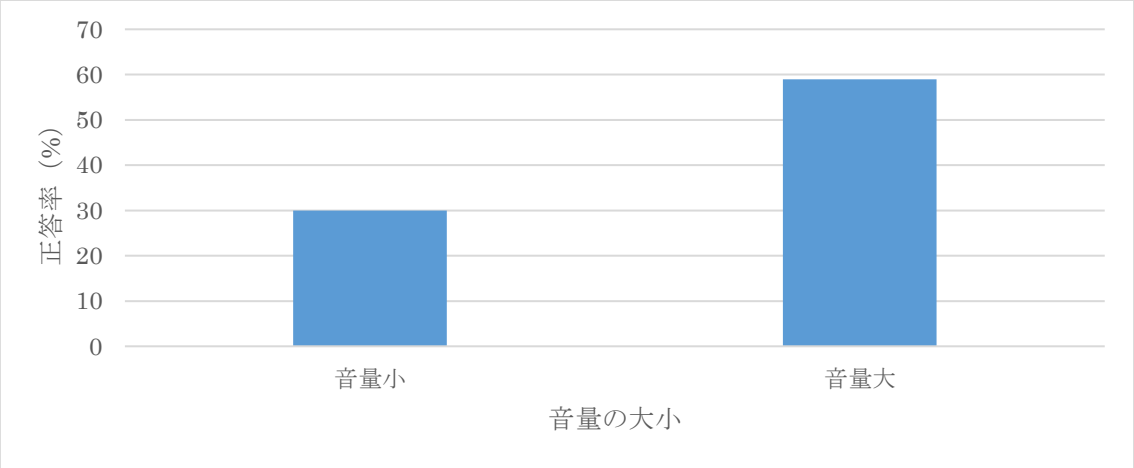

図 7　音量別正答率

4. おわりに

本研究では環境雑音を用いた音声 CAPTCHA を提案した．提案方法においてテキストを長くすることにより HAR を高め，MAR を低くなることが明らかになった．しかし，提案 CAPTCHA には自動生成に課題があり，環境雑音を自動的に生成することが難しい．今後は更に HAR を高めて MAR を低くするとともに，素材からの自動生成を目指すことを目標とする．

謝辞

本研究においてご指導をいただいた菊池浩明教授，実験に協力していただい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学部生の皆さま，明治大学先端数理科学研究科先端メディアサイエンス専攻菊池研究室の院生の皆さまに感謝いたします．

参考文献

[1] 藤田真治，池谷勇樹，可児潤也，西垣正勝，“非現実画像 CAPTCHA：常識からの逸脱を利用した 3 DCG 画像 CAPTCHA”，情報処理学会論文誌，Vol.56，No.12，pp.2324-2336，2015.

[2] 滋野莉子，山田道洋，山口通智，菊池浩明，坂本真樹：“オノマトペ CAPTCHA の開発と評価”，マルチメディア・分散・協調とモバイルシンポジウム，pp.1778-1785，2017

[3] 滋野莉子，山田道洋，菊池浩明，坂本真樹：“オノマトペ CAPTCHA の開発と評価：日英の比較”，第 22 回曖昧な気持ちに挑むワークショップ，pp84-89，2017

[4] 西本卓也，西亀健太，嵯峨山茂樹，福岡千尋，渡邊隆行，“音声 CAPTCHA のための音韻修復効果の検討”，聴覚研究会資料，Vol.38，No.6，pp.639-644，2008.

[5] 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識方式”，コンピュータセキュリティシンポジウム 2017 論文集，pp.966-971，2017.

[6] 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，pp.363-370，2016.

[7] Google Is Creating A Newer And Better CAPTCHA System（https://www.androidheadlines.com/2014/12/google-creating-newer-better-captcha-system.html）.

[8] ReCAPTCHA demo（https://www.google.com/recaptcha/api2/demo）.

付録 A　近赤外線分光法 NIRS を用いたストレスの強さの定量化

A.1. はじめに

A.1.1. 研究背景

ボットによる Web サービスの不正利用を防ぐために CAPTCHA が使用されている. しかし, 近年では主流の文字画像認証型の CAPTCHA が OCR を導入したボットに突破されている．従って，人間には易しく，機械には困難であり，問題の自動生成が可能である新しい CAPTCHA を作る必要がある．だが，CAPTCHA が複雑化すると人間に過度なストレスを与えていることが懸念される.

そこで，本研究では近赤外線分光法 NIRS(Near InfraRed Spectroscopy)を用いて脳内ヘモグロビンの酸素含有量を見ることによって，ストレス度数の指標を作成することを目的とする.

A.1.2. 研究目的

本研究は,CAPTCHA によるストレスを定量化することを目的とする.

A.1.3. 近赤外線分光法 NIRS

00mm から 950mm の近赤外線は生体組織を破壊することなく透過できる．その性質を用いて脳内の酸素濃度や血中ヘモグロビン濃度の変化を見ることができる.

本研究では，ダイナセンス社が製造している近赤外線分光法を用いたヘッドマウント型の組織酸素モニタ装置である PocketNIRSHM を使用する．（以下，NIRS と呼称する.）

A.2. 実験

A.2.1. 実験概要

明治大学総合数理学部に所属している 12 名の学生を被験者とし，NIRS を付けている状態で皿に入っている 10,20,30 個のビーズを箸で他の皿に移すという作業を行ってもらった．このビーズ移動作業のイメージを図 8 に示す.

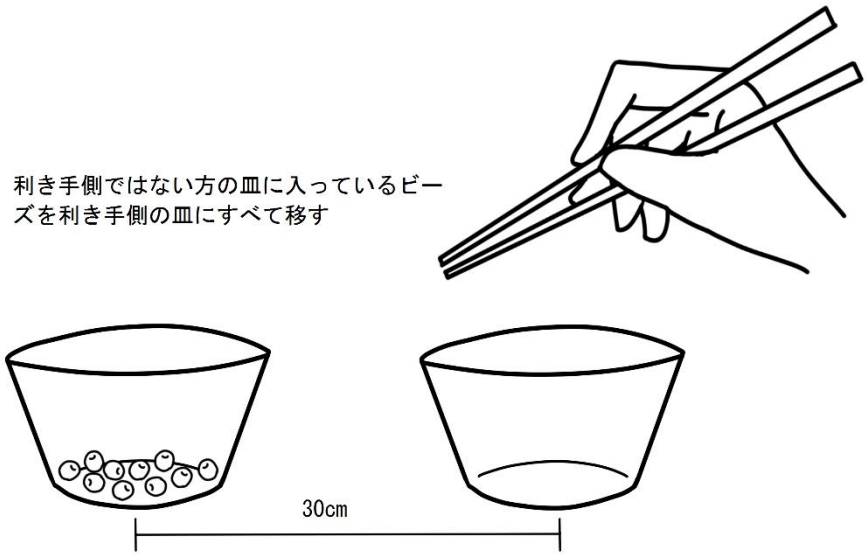

図 8　ビーズ移動作業のイメージ

A.2.2.　　データ形式

NIRS によって収集したデータの例を図 9 に示す．図 9 はビーズ移動作業中の観測結果である．図 2 の 3 つある測定値は，上から順に酸素を有するヘモグロビンの量，酸素を有するヘモグロビンと有さないヘモグロビンの合計量，酸素を有さないヘモグロビンの量，その合計量である．縦にひいてある線はイベントフラグであり，作業中とそうでない時の仕切りである．

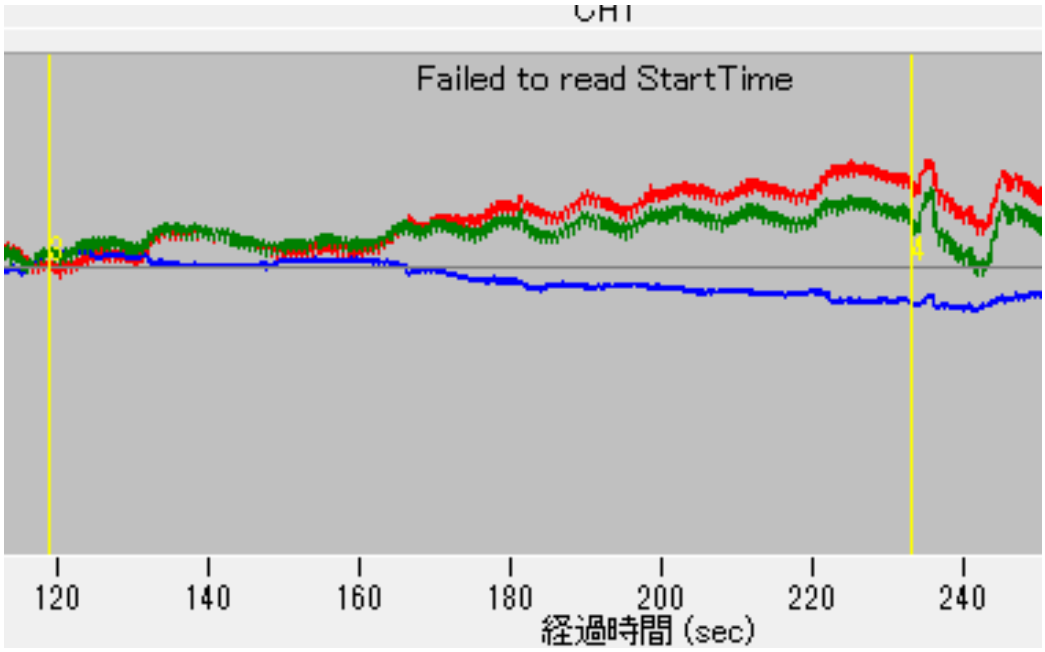

図 9　収集したデータの例（一部切り取り）

A.2.3.　　ストレス指標の提案

ビーズ 10 個単位を 1 つのレベルと定める．指標に使用する観測データを表 9 に示す．本実験ではビーズ 30 までの実験のため観測データでの最大レベルは 3 となっている．ストレス度数，レベルの二つの要素からなる．本研究ではストレス度数を NIRS によって取得した数値（NIRS によって取得する数値に明確な単位は存在しないためここでは nir(ニル)と定義する.）について，$x_1,\ldots,x_n$の測定値から

$$\text{nir} = \max(x_i) - \text{平均値}x_i$$

と定める．

表 9　観測データ（一部）

ストレス度数[nir]	レベル[Level]
-0.00292	1
0.028901	2
0.06986	3
-0.00666	1
0.024867	2
0.051864	3
0.020969	1
0.030183	2
-0.00126	3
0.036467	1

A.2.4.　　実験結果

レベル毎のストレス度数を R で線形回帰分析をした．結果を図 10，表 10 に示す．個人別に様々な傾きではあるが被験者それぞれの傾きから算出した結果，84.6%は正の傾きであった．線形回帰の傾きの平均値は 0.021949[nir/Level]であった．

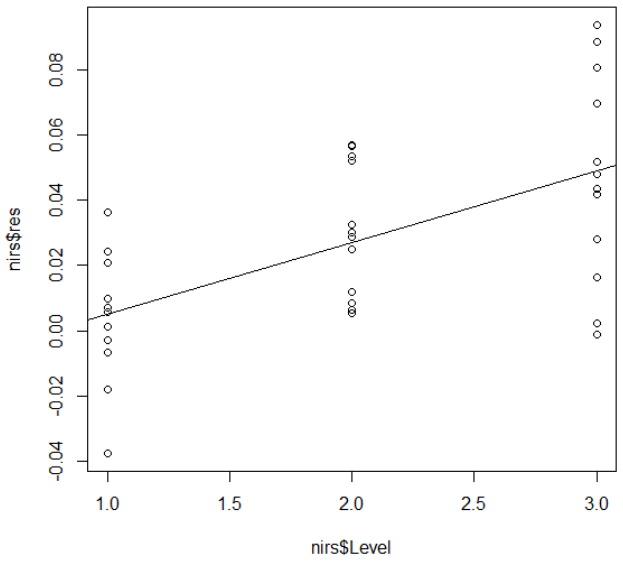

図 10　ストレス指標

表 10　ストレス度数の統計値

	ストレス度数[nir/Level]
平均の傾き	0.0429
傾きの最大値	0.06587
傾きの最小値	-0.01111
傾きの標準偏差	0.023146
全体の傾き	0.021949

A.2.5. 考察

本実験では大きな 3 つの問題点があった．

⑴ NIRS を使うことによるもので，キャリブレーションにある．NIRS はキャリブレーションした後に測定するため，被験者のもともとの状態に大きく左右される．また，相対的な数値しか取れないため，定量的な値を算出するのには工夫が必要である．

⑵ 個人差やアクシデントである．箸を使うことに対する得手不得手の個人差やビーズを落とすといった，レベルとは直接的には関係のない不慮のアクシデントが生じた．

⑶ ストレス度の定義である．本実験ではストレス度数としてレベル毎の最大値から作業中のみの平均値を引いたもの用いていたが，正当性が不明であった．

A.3. おわりに

本実験により，ストレス度数の指標として 10 個ビーズの作業に相当するレベル毎のストレス度数の増加分 0.021949 を算出した．これをもとに様々な行動に対するストレス度をレベル毎に割り出すことが可能になった．

今後は，この指標を用いて既存の CAPTCHA やオリジナルの CAPTCHA のストレス度数の評価を行い，人間に対してストレスがあるかどうかを明らかにすることを課題とする．