環境雑音を用いた音声 CAPTCHA の認識実験

二谷太郎†
明治大学総合数理学部　先端メディアサイエンス学科　菊池研究室†

1　はじめに

ボットによる Web 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)が使用されている. 図 1 に主流である文字画像判別型 CAPTCHA の例を示す. しかし,近年ではこ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,CAPTCHA の必要条件である,

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,

に基づいた新しい CAPTCHA の開発が必要とされている [1].

図 1　文字画像判別型 CAPTCHA の例

[2][3][4]では音声を用いた方式が提案されている. 既存の音声を用いた CAPTCHA は非常に難易度が高いものが多く,CAPTCHA の必要条件を満たせなくなっている.

そこで,本研究では,環境雑音(カフェの作業音や駅前の雑踏など)をノイズとして使った音声 CAPTCHA を提案する. 人間は従来から様々な環境雑音の中で会話することに慣れているため回答するのは容易であり,機械にとっては雑音に紛れている音声の判別が困難であると予想される.

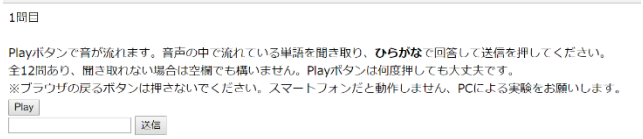

図 2　実験ページ

2　環境雑音を用いた音声 CAPTCHA

2.1　提案方式

以下に提案方式の構成を示す.

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(2～8 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(音程は平均よりやや高めのもの)を使用した.
3. 雑音と単語を重ね,再生する.

図 2 に実際の実験ページを示す.

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 を回答文字数が 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. 本実験では問題数に対して正しく解答できた割合を HAR とする.

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 の実験結果を表 2 に,文字数別の HAR を表 3 に示す.

表 2 より,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 文字の問いが最も HAR が高かった. 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.

† Taro Futatsuya, Department of Frontier Media Science, School of Interdisciplinary Mathematical Science, Meiji University, Kikuchi Laboratory.

表 2　男女別,年齢別の HAR

		N	平均	標準偏差	最大値	最小値
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	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 問のノイズと合成した CAPTCHA と単語の音声データのみの 12 問の計 24 問を Siri と Google の音声入力にそれぞれ 10 回ずつ行った.本実験では試行回数に対する正しく認識できた割合を MAR とする.

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 の音声入力の認識結果の例を表 5 に示す.単語の隣にある()の中の数字は認識した回数を示す.

表 5 より,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.テキストが長いほど一語を複数文節からなる文と誤認をする可能性が増え,機械には認識されにくいことが分かった.

表 5　Google の音声入力の認識結果の例

もず	認識しない(9),マック(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),一月(1)
かんこんそうさい	おそ松さん 3(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)
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,長いテキストの問題は HAR が高く,MAR が低いことが明らかになった.この傾向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　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で 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 には自動生成に問題があり,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 を低くしつつ,素材からの自動生成を目指すことを目標とする.

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 3DCG 画像 CAPTCHA”, 情報処理学会論文誌, Vol.56 , No.12 , pp.2324-2336, 2015.

[2] 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.

[3] 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.

[4] 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.