環境雑音を用いた音声CAPTCHAの認識実験

明治大学総合数理学部

菊池研4年　二谷太郎

CAPTCHA

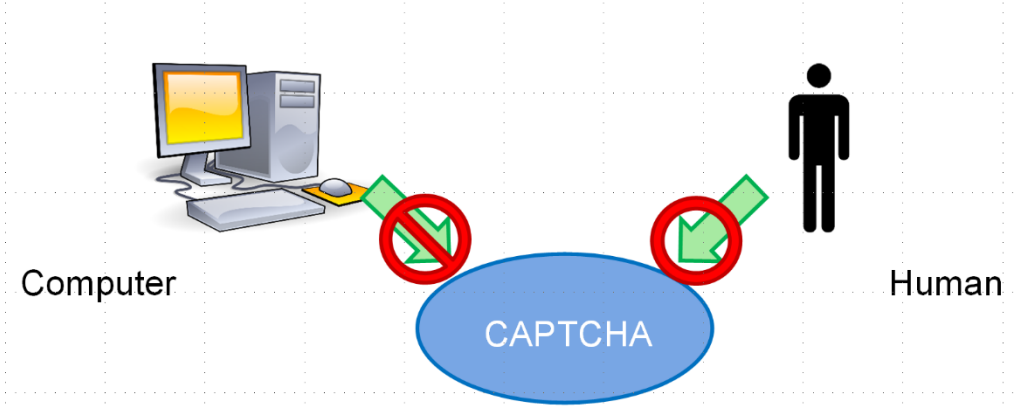
- CAPTCHA

(Completely Automated Public Turing test to tell Computers and Humans Apart)

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

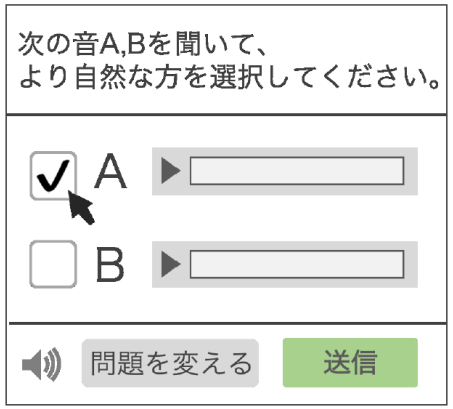
- CAPTCHAの必要条件
  - 人間にとって解くのが容易である
  - 機械にとって解くことが困難である
  - 問題の自動生成が可能である

- 例）文字画像判別型CAPTCHA



先行研究

- Google reCAPTCHA
  ノイズに混じった0から9の数字5文字を聞き取らせ，利用者に解答させる方式
  - 問題点
    - 聞き取りが難しい
    - 回答に時間がかかる
- 混合された環境音の聞き取りに基づく認証方式（古賀・佐藤）
  ある一つの環境音に対して，関係のある音を組み合わせた混合音A と関係のない音を組み合わせた混合音B の二つのうち，どちらが自然であるか（状況としてふさわしいか）を利用者に解答させる方式
  - 問題点
    - 問題の定義があいまい
    - ブルートフォース攻撃に非常に弱い



環境雑音を用いた音声CAPTCHAの提案

- 人間が普段様々な環境雑音（カフェの作業音や駅前の雑踏など）の中で会話することができることに着目し、環境雑音をノイズとして使った音声CAPTCHAが有用であると考えた

- 提案方式は以下の通りである

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，切り取る

② 日本語の単語をOpen JTalkを使用して音声合成する

③ 環境雑音と単語の音声を重ね，再生する

HARとMAR

- HAR(Human Acceptance Rate)：人間を正しく受け入れる確率
  - 本研究では実験における正解率をHARとする
- MAR(Machine Acceptance Rate)：機械を誤って受け入れる確率
  - 本研究では音声認識実験における正しい認識の割合をMARとする

		真	
		人間	機械
判別	人間	A	B
	機械	C	D

- HAR（高いほどよい）

$$\frac{A}{A+C}$$

- MAR（低いほどよい）

$$\frac{D}{B+D}$$

実験

- 実験1　HAR
  - HARを求め提案方式の精度を明らかにするのを目的とする
  - 環境雑音は，2.5秒程度の長さに切り取った
  - 全12問，解答が2,4,6,8文字のものを各3問ずつ，よく聞く単語や聞きなじみのない単語をそれぞれの文字数のものに入れた
  - 使った音声は，女性ボイス（平均よりやや高めのものを使った）
  - 学部生，院生を対象として44名にウェブ上で実験を行った

- 実験2　MAR
  - MARを求め提案方式の精度を明らかにすることを目的とする
  - 実験1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の音声入力にそれぞれ10回ずつ行った

デモ

1問目

Playボタンで音が流れます。音声の中で流れている単語を聞き取り、**ひらがな**で回答して送信を押してください。
全12問あり、聞き取れない場合は空欄でも構いません。Playボタンは何度押しても大丈夫です。
※ブラウザの戻るボタンは押さないでください。スマートフォンだと動作しません、PCによる実験をお願いします。

Play

送信

答え：おはよう

答え：とくがわつなよし

実験1　実験結果

性別，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

文字数別のHAR

文字数	HAR
2	0.772
4	0.659
6	0.598
8	0.810

実験2　実験結果

認識システムによるMAR

		認識システム（MAR）			
		Siri（単語単体）	Siri（CAPTCHA）	Google（単語単体）	Google（CAPTCHA）
文字数	2	0.000	0.000	0.333	0.067
	4	0.800	0.000	0.500	0.133
	6	0.333	0.000	0.467	0.033
	8	0.667	0.000	1.000	0.000
計		0.450	0.000	0.575	0.058

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)

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)

先行研究との比較

CAPTCHA	ブルートフォース耐性	HAR
提案方式（8文字のとき）	1/71^8 （約646兆分の1）	0.810
先行研究（古賀・佐藤）	1/2	0.746
reCAPTCHA	1/100000	0.872

CAPTCHA	問題数	HAR
提案方式（8文字のとき）	1	0.810
先行研究（古賀・佐藤）	49.2	5.48E-7
reCAPTCHA	2.96	0.667

ブルートフォース攻撃耐性に合わせてHARに補正を掛けてみる

まとめ

- 提案方式において，文字数を8文字程度まで長くすることにより，HARが低い文字数のものと比べて0.212高くなり，機械が著しい誤認をしてMARが低くなることが分かった

- 自動作問に問題があり，環境雑音を自動生成するのが難しいという点があるが，素材からの自動作問が今後の課題である