

# 音声 CAPTCHA のための音韻修復効果の検討

西本 卓也<sup>†</sup> 西亀 健太<sup>†</sup> 嵯峨山茂樹<sup>†</sup> 福岡 千尋<sup>††</sup> 渡辺 隆行<sup>††</sup>

<sup>†</sup> 東京大学大学院 情報理工学系研究科

〒 113-8656 東京都文京区本郷 7-3-1

<sup>††</sup> 東京女子大学 現代文化学部 コミュニケーション学科

〒 167-8585 東京都杉並区善福寺 2-6-1

あらまし 対象者が人間であるか機械であるかを判別する CAPTCHA システムを音声によって実現するための基礎的な検討を行った。このようなシステムは視覚障害者のウェブ利用にとって重要であるが、破られにくさや使いにくさなどの課題がある。まず我々は CAPTCHA システムの設計方針を「人間と機械の能力差を作り出す」という観点から体系化した。次に音韻修復効果がこの目的に利用できるという仮説に基づいて、人間による聴取実験と HMM による音声認識実験による予備的検討を行った。録音音声を用いた実験の結果、マスキング領域の割合を適切に設定すれば人間の了解度が機械による音声認識性能を大きく上回る可能性が示唆された。合成音声を用いた場合にも同様の効果が得られることが期待される。

キーワード インターネット, セキュリティ, 視覚障害, 音声認識, 音声合成

## Investigation of Phonemic Restoration for Speech CAPTCHA Systems

Takuya NISHIMOTO<sup>†</sup>, Kenta NISHIKI<sup>†</sup>, Shigeki SAGAYAMA<sup>†</sup>, Chihiro FUKUOKA<sup>††</sup>, and

Takayuki WATANABE<sup>††</sup>

<sup>†</sup> Graduate School of Information Science and Technology, The University of Tokyo

<sup>††</sup> Department of Communication, Tokyo Woman's Christian University

**Abstract** We carried out preliminary experiments to realize a speech-based CAPTCHA system to distinguish between software agents and human beings. Such systems are especially important for persons with visual disability. Current technologies are, however, vulnerable and difficult to use. As the first step, we investigated the design principles of CAPTCHA systems from viewpoint of creating the performance gap between the machines and human beings. Based on hypothesis that the Phonemic Restoration effects are useful for this purpose, we carried out experiments of intelligibility tests by human beings and speech recognition tests with HMM-based systems. As the results with recorded human voice, it was suggested that human intelligibility exceeds speech recognition performance under the condition of appropriate ratio of masking area. It is expected that similar effects are observed when we use synthesized speech.

**Keywords** Internet, Security, Visual Impairment, Speech Recognition, Speech Synthesis

### 1. はじめに

近年, インターネットにおいて人間を装ったソフトウェアロボットが掲示板, WIKI サイト, ブログのコメントやトラックバックなどにおいて広告目的の無関係な書き込みを大量に行ったり, 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, といった状況が増えており, インターネットサービスの円滑な運営が妨げられている. これを防ぐために多くのサイトで CAPTCHA ( Completely Automated Public Turing test to tell Computers and Humans Apart ) と呼ばれるシステムが用いられるようになってきた. これは対象者が人間であるか機械 ( コンピュータ ) であるかを判別するテスト

トのことである。

多くの CAPTCHA システムでは画像データに含まれる文字を人間が読み取る手法が用いられている。字形を歪めた文字列や文字の一部が隠されているものを機械的に生成し、これを HTML コンテンツの中で提示する。いわゆる機械による自動文字認識が容易に行えないようになっている。画像に描かれている文字列を読み取れたのであれば、その登録者は人間であろうと判断し、アカウントの取得を認めるようになっている。しかし、画像を使用する CAPTCHA は視覚障害者のインターネットの利用を妨げているという側面がある<sup>(注1)</sup>。そこで視覚障害者を対象として音声 CAPTCHA による認証を行うポータルサイトも登場している<sup>(注2)</sup>。

一方で、CAPTCHA を破る技術が発達し、実際にシステムが破られているという指摘もある<sup>(注3)</sup>。文字や数字を読み取らせるのではなく、より複雑な課題を用いた CAPTCHA システムについての提案もあり、例えば 3-D CG による人物画像を読み取って人物の体の部位などを回答させる 3-D CAPTCHA<sup>(注4)</sup>などが考案されている。

CAPTCHA の技術は機械によるパターン認識技術の実情を踏まえた「破られにくい」ものでなくてはならない。また、Web アクセシビリティ [1] の観点からは、画像による CAPTCHA と同じ程度に破られにくく使いやすい音声 CAPTCHA を実現することは重要である。

本報告ではまず CAPTCHA システムの設計方針を「人間と機械の能力差を作り出す」という観点から体系化する。次に音韻修復効果がこの目的に利用できるという仮説に基づいて、人間による聴取と HMM による音声認識に関して行った予備実験について述べる。最後にテキスト音声合成を用いる手法の提案と、今後の課題について述べる。

## 2. 音声 CAPTCHA の設計方針

### 2.1 基本的な考え方

人間の視覚においては、意味のある情報をトップダウン的に読み取ることができる、といった特性がある。人間はこの特性を利用して、断片的であったりノイズが混じっていたり大きく歪んでいる文字を読み取ることができる。しかし機械による自動画像認識は一般的にこのような条件下では性能が劣化する。このように、人間の知覚特性とパターン認識技術の双方の特性を有効に利用するのが、CAPTCHA システムの設計における基本的な考え方となる。これは画像に限らず、音声による CAPTCHA

でも同じである。

### 2.2 破られにくさの考慮

前述した Google や Microsoft の音声 CAPTCHA システムでは、例えば数字を読み上げる音声に雑音を重畳したものが用いられている。その際、残響を加える、日本語で読み上げた数字に対して他の言語の音声を断片化して重畳する、といったことが行われる。人間の聴覚に対しては「カクテルパーティ効果」のような現象を期待し、自動音声認識に対しては性能を大きく劣化させる妨害の効果を期待している。

現在の自動音声認識の主流である HMM (Hidden Markov Models) を用いた手法は、残響や雑音で劣化した音声であっても教師あり学習によって性能を向上できる。ラベル付き学習データを作成する手間に見合った性能向上が見込めないような状況を作り出すことが「破られにくさ」につながる。

さらに、提示される音声のバリエーションが少ないと、大きく劣化した音声であってもテンプレートマッチングによって破られる可能性がある。例えば Google の音声 CAPTCHA を簡単な手法で破ったとするブログ記事<sup>(注5)</sup>では、同一の数字音声パターンが繰り返し使用されていることが脆弱性の原因である、と指摘している。

### 2.3 使いやすさの考慮

誰でも知っている身近な情報を使う CAPTCHA としては Microsoft Research の Asirra<sup>(注6)</sup>が挙げられる。これは膨大なペットの写真データベースを利用して、画像が犬と猫のいずれであるかを利用者に回答させる手法である。この手法は、犬と猫の画像の識別は誰にでもできる、という妥当な前提に基づいている。また膨大なバリエーションが利用可能であるため自動パターン認識が困難になる。しかし個々の課題のチャンスレベルが 50%であるため、安全な認証を行うためには複数の質問を行う必要がある。

アクセシビリティの観点からは Holman らの提案 [2] も興味深い。これはサイレンや鳥などの身近なオブジェクトを視覚と聴覚の両方で提示し、そのいずれによっても回答可能とすることで、アクセシブルで多言語化が容易な CAPTCHA を実現する。回答は選択肢によって行う。しかしこの手法では多くのバリエーションを提供することができなければ簡単に破られてしまう。

インターネット利用者の認証に用いるという観点からは、「誰もが知っている知識」を前提にしなくてはならない。例えば「有名人の顔写真を選ぶ」「難しい数学の問題を解かせる」といった個人の知識や知的能力に依存する手法は好ましくない。文字や数字の読み取りは誰もが利用できる代表的な知識である。

また、実現可能性という観点からは、キーボードで回答を入力できる、あるいは簡単な選択肢操作で回答できる課題であることも望まれる。総当たりやランダムでソフトウェアロボットが攻撃することを防ぐためには、チャンスレベルが低くなる課題設

(注1): Inaccessibility of CAPTCHA, Alternatives to Visual Turing Tests on the Web, W3C Working Group Note 23 November 2005. <http://www.w3.org/TR/turingtest/>

(注2): 例えば Google 「Google アカウントを作成」、Microsoft 「Windows Live ID サインアップ」など

(注3): CAPTCHA 認証は「終わった」技術なのか 有効性を疑問視する専門家たち <http://www.computerworld.jp/topics/vs/115709.html>

(注4): <http://spamfizzle.com/CAPTCHA.aspx>

(注5): <http://blog.wintercore.com/?p=11>

(注6): <http://research.microsoft.com/asirra/>

定も必要である．このような観点からも，やはり数字や文字の読み取りは，さまざまな要求を満たしやすい課題と言える．

心的負荷や所要時間の観点から「使いやすさ」に考慮することも重要である．視覚障害者がスクリーンリーダーや音声ブラウザでインターネットを利用する際には，キーボードと音声出力だけが利用可能であり，視覚情報やマウス操作に頼ることができない．また多くの情報を記憶しながら操作をしなくてはならない．このような作業は煩雑であり訓練を要する．「破られにくさ」を重視するために音声 CAPTCHA では桁数の多い数字を聞き取って回答することが行われるが，記憶に伴う心的負荷も増大し，所要時間においても不利である．できるだけ短時間で多くの情報を聞き取ることができ，その際の利用者の心的負荷が小さく，しかも破られにくい，という手法が望まれる．

#### 2.4 混合法と削除法

現在の音声 CAPTCHA システムの主流は，人間に聞き取らせたい音声に残響や雑音を重畳する，という手法である．この手法をここでは「混合法」と呼ぶ．

混合法は，音声認識技術が雑音や残響に弱いという性質と，人間が複数の音の重なりの中から興味を持つ音だけを聞き取ることができる，という能力を利用している．現時点ではこの手法の有効性は認められるが，将来的に破られにくくしていくために，どのような雑音をどのように混合すれば破られにくくできるか，といった調整は試行錯誤的になり，パラメトリックに難易度を制御しにくい．

一方で，人間の聴覚に関してはさまざまな興味深い特性が知られており，CAPTCHA に応用可能なものもあると考えられる．その中のひとつとして，今回我々は「音声の断片から全体を推測して聞き取る能力」に着目し，「原音声からより多くの部分を削除するほど困難な課題になる」と考えた．このような手法をここでは「削除法」と呼ぶ．

混合法と削除法は相反する考え方ではなく，併用が可能だと考えられる．我々が削除法に着目する理由は，難易度をパラメトリックに制御しやすいという点にある．つまり，原音声をより多く残せばそれだけ容易になり，原音声を残す割合を少なくすればそれだけ困難になる．このような制御の容易さは混合法にはない特長である．

#### 2.5 認識能力ギャップのモデル

本研究では，削除法に基づく音声 CAPTCHA において，図 1 の設計モデルの妥当性を検証し，基礎的なデータを得ることを目標とする．CAPTCHA 課題としての適切さは，図中の矢印で示す「機械と人間の認識能力のギャップ」の大きさによって示される．このギャップが大きく条件を見つかることができるかどうか，削除法の有効性の判断基準となる．

### 3. 実験方法

#### 3.1 音韻修復効果の考慮

一般的に音声の一部を削除して聴取を行うと，その了解度は

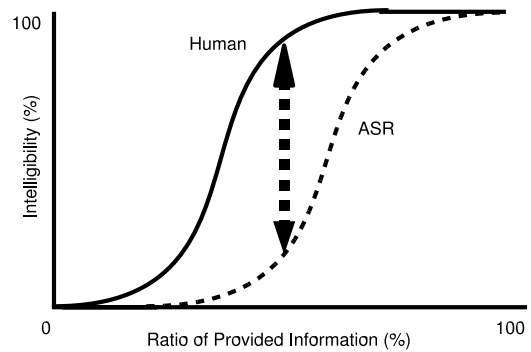


図 1 音声 CAPTCHA 設計の概念モデル

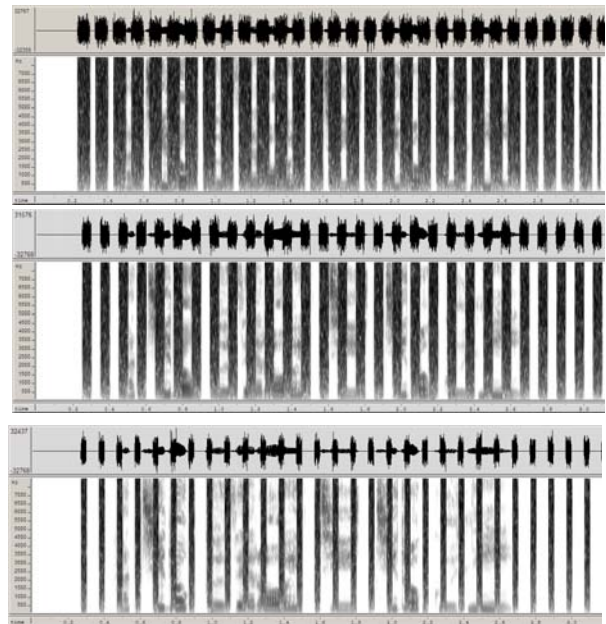


図 2 使用する音声ファイルの例．7 桁の数字読み上げ音声に対して，上から順に 30%, 50%, 70% の割合で原音声を残し，ホワイトノイズを埋めたもの．

低下する．しかしただ削除するのでなく，削除した部分に別の音を挿入すると，音声の聴取が容易になることが知られている．これが「音韻修復」である．音韻修復については様々な研究が先行しており，代表的なものとしては Miller and Licklider (1950) による「垣根効果」の報告が挙げられる [3] ~ [5] ．

削除法における人間の聴覚能力を高めることは，前述した認識能力ギャップの拡大に貢献する可能性が高い．そこで我々は，音声を周期的に決まった割合で削除し，削除された区間をホワイトノイズで埋めることにした．使用する音声の波形とスペクトログラムの例を図 2 に示す．原音声を残した割合が少ないほど人間にとって困難な課題となるが，同時に音声認識システムに対しても不利な課題となることを期待している．

#### 3.2 使用する音声データ

今回我々は残響下音声認識評価環境である CENSREC-4 [6] に含まれる数字読み上げ音声データベースを使用した．これは，

1~7桁の数字読み上げという課題が音声 CAPTCHA システムを模擬するうえで適切だったこと、男女の別を含む多数の話者によって読み上げられた音声であり、破られにくい CAPTCHA という観点からも妥当と考えられたからである。

ただし将来は、音声合成や声質変換などの手法を用いて、テンプレートマッチングや教師あり学習などによる自動認識がより困難な方法を検討する必要がある。

CENSREC-4 に含まれるクリーン発話音声 (16kHz 16bit Mono) を用いて、100ms 周期で原音声を削除し、ホワイトノイズを埋め込んだ。条件としては 70%、50%、30% の3種類のデータを作成した。例えば 70% 条件の場合は原音声の 100ms ごとに 70ms を残し、30ms を雑音区間とした。音声および雑音の立ち上がりりと立ち下がりには 10ms の幅を持たせて、直線エンベロープでクロスフェードする処理を施した。周期を 100ms としたことで、このようなマスキングを行っても数字全体の欠落が起りにくくなっている。

### 3.3 HTK による音声認識実験

CENSREC-4 の実験系に準拠して音声認識実験を行った。

この実験の目的は、ホワイトノイズによってマスキングされた音声ファイルとその正解ラベルが大量に与えられている、という前提で、既存の音声認識エンジンをそのまま使用した場合の性能を知ることである。大量の教師あり学習データを揃えることは実際には容易ではない。従って現実の利用状況は「破る側」にとってより不利な状況だと考えられる。その反面、ホワイトノイズの区間を判別する前処理を行うなど、削除法に特化した「破り方」を用いることで、「破る側」はさらに有利になる可能性もある。今回の実験はこのような諸要因を考慮するための準備段階という位置づけになる。

音声認識には HTK<sup>(注7)</sup> を用いた。3 条件のそれぞれについて、学習データは 8440 発話、評価データは 1001 発話である。両者は発話そのものも話者も重複していない。音響モデルには単語 HMM を用いた。数字の 0 に「ぜろ」「まる」の 2 種類の読み上げ方があるため 11 種類のモデルを構築した。各単語は 18 状態 20 混合である。評価は発話全体 (全ての桁の数字) の一致によって判断した。

### 3.4 人間による聴取実験

前述した評価データの一部を用いて人間による聴取実験を行った。

まず回答者の記憶の負荷や入力操作の容易さを考慮して、評価データのサブセットを作成した。1001 個の評価データから 3 桁、4 桁、5 桁の数字が 25 個ずつ含まれる 75 個の課題セットを 3 組作成した。この際、数字 0 については「ぜろ」の読み方のみが含まれるようにした。課題セット間で各数字の出現頻度の偏りができるだけ少なくなるように配慮した。課題セットごとに出現順序はランダムに並べ替えた。

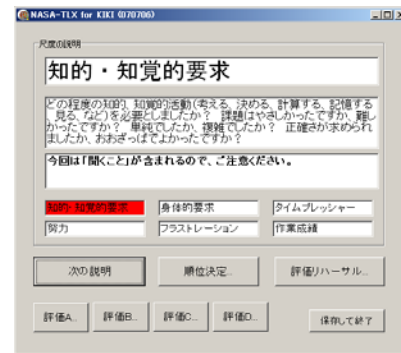


図 3 NASA-TLX プログラムのメインメニューと尺度の説明。

被験者は大学生 (女性) 17 人で、全員が日本語を母語とする健聴者である。

### 3.5 NASA-TLX による心的負荷の評価

聴取実験に合わせて心的負荷の主観評価を行った。これは課題が困難になるにつれて心的負荷が高くなるのではないかと、という仮説を検証するためである。

NASA-TLX (Task Load Index) [7], [8] は以下の 6 つの下位尺度によって心的負荷を評価する手法である。

- 知的・知覚的要求 (小さい / 大きい)
- 身体的要求 (小さい / 大きい)
- タイムプレッシャー (弱い / 強い)
- 努力 (少ない / 多い)
- フラストレーション (低い / 高い)
- 作業成績の悪さ (良い / 悪い)

被験者は負担度評価の対象となる作業を遂行する前に、下位尺度の重要度を評価する。作業のあとで被験者は 6 つの尺度それぞれに対する評定を (本研究では 0~100 のスクロールバー操作によって) 行う。重要度の評定に基づいて、下位尺度の評定値の mean weighted workload score (加重平均作業負荷得点: WWL 得点) を求める。本実験では重要度が最上位と評価された尺度から順に重みを 6~1 とした。

実験には我々が作成した Windows 用ソフトウェア Kiki-WWL を使用した。図 3 はメニューと各尺度の説明、図 4 は尺度の順序づけの画面である。課題間の値の大小関係を意識した評定が行われるように考慮している。尺度の説明においては「作業成績 = すべて正しく聞き取って入力することが目標」「知的・知覚的要求 = 聞くことを含む」「身体的要求 = 聞くこと・入力することを含む」などを補足した。

### 3.6 聴取実験の手順

被験者を 3 群に分けて聴取実験を行った。

1 人 1 台のノート PC (Microsoft Windows XP) とヘッドフォン (Panasonic RP-H750-S) を使用した。被験者自身に音声提示および Kiki-WWL のソフトウェアを操作させて、10 秒間隔で再生される音声を 1 つずつ聴取させ、聞こえた数字をコンピュータのキーボードから入力させた。

(注7): <http://htk.eng.cam.ac.uk/>



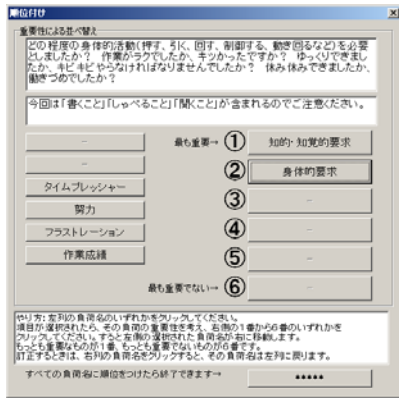


図 4 NASA-TLX プログラムにおける尺度の重み付け画面。

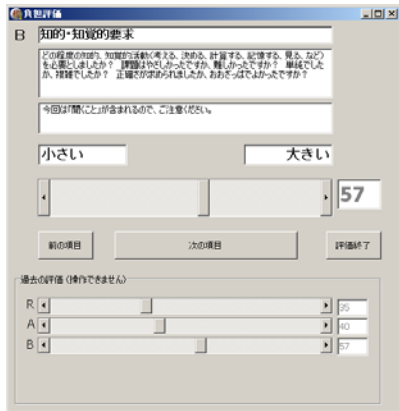


図 5 NASA-TLX プログラムにおける各尺度の負荷評価の画面。

表 1 聴取実験の構成

Group	Trial 1	Trial 2	Trial 3
G1	V1-70%	V2-50%	V3-30%
G2	V1-30%	V2-70%	V3-50%
G3	V1-50%	V2-30%	V3-70%

最初に音量の確認を兼ねたりハーサルのためにマスキングされた 2 桁の数字音声を 10 問呈示し、本番同様に回答させた。次に、音声ガイドと文字表示で下位尺度の説明を行い、下位尺度の順位付けをさせ、リハーサル課題に関する各尺度の負荷評価を行わせた。その後 3 回の試行について、(1) 75 問の聴取を行わせて (所要時間は約 13 分)、(2) 6 つの尺度の負荷評価を行わせ、(3) 疲労の影響を回避するために 5 分間の休憩を取る、という手順を繰り返した。被験者は表 1 のように分けた。V1-V3 は課題セット (語彙) が異なることを、後続する数字は原音声を残した割合 (値が小さいほど多くマスキングしている) 意味している。

NASA-TLX で得られた重み付き心的負荷 (WWL) の値は、被験者ごとに平均 50、標準偏差 10 となるように正規化した。

## 4. 実験結果と考察

### 4.1 HTK による音声認識の結果

HTK による音声認識結果を表 2 に示す。評価データと同じよ

表 2 HTK による文認識率 (%)

原音声の割合	30%	50%	70%
文認識率	40.46	79.02	89.41

表 3 聴取実験における文了解度 (%)

原音声の割合	30%	50%	70%
平均	84.16	97.41	98.98
標準偏差	6.51	2.72	1.97

表 4 聴取実験における正規化心的負荷

原音声の割合	30%	50%	70%
平均	60.55	45.21	44.24
標準偏差	5.45	7.48	6.85

うにマスキングされた音声を用いて音響モデルを学習することで、原音声 50% の条件であれば約 80% の認識率が得られる。一方で 30% 条件においては約 40% と大きく認識率が低下した。学習データとしてマスキング音声を大量に集めても、削除された区間が大きい場合は、従来の音声認識手法では認識が困難になると考えられる。特に今回用いたデータにおいて話者のバリエーションが非常に多いこともこのような結果をもたらした原因であろう。

### 4.2 聴取実験における了解度

聴取実験における了解度の被験者 17 人についての平均および標準偏差を表 3 に示す。両側分布  $t$  検定の結果、30%-50% および 30%-70% の各群間において 1% 水準で有意差があった。原音声の割合が小さくなるほど了解度は低下するが、30% の条件においても音声認識実験と異なり 80% 以上の了解度が得られた。このことから、削除法と音韻修復効果を用いることで、人間と機械の認識能力の大きなギャップを作り出せる可能性が示唆された。

### 4.3 聴取実験における心的負荷

被験者ごとに正規化した NASA-TLX の値 (Normalized-WWL) の被験者 17 人についての平均および標準偏差を表 4 に示す。両側分布  $t$  検定の結果、30%-50% および 30%-70% の各群間において 1% 水準で有意差があった。了解度が有意に下がる条件では心的負荷の有意な増加も確認された。心的負荷の測定について本手法の妥当性が示されたと言える。

## 5. 音声合成の利用の検討

現在、CENSREC-4 の録音音声の代わりに、音声対話技術コンソーシアム (ISTC) <sup>(注8)</sup> のもとで開発されている HMM 音声合成エンジン GalateaTalk <sup>(注9)</sup> の合成音声を用いることを検討している。100ms 周期でマスキングした音声による予備的な検討では、合成音声でも録音音声と同様に音韻修復効果が得られるという見通しを得ている。

合成音声を用いることで音声のバリエーションを増やすこと

(注8): <http://www.astem.or.jp/istc/>

(注9): <http://sourceforge.jp/projects/galateatalk/>

が容易になる。統計モデルの学習を困難にできれば、音声認識によって破られることを妨害できる。しかしそのためには、人間にとっての了解度を落とさずに発話様式や声質などをできるだけ多様化する手法が必要となる。

一方で、音声合成システムを用いることで、韻律や発話速度・リズムなどの要因がどのように音韻修復に影響をするのか、といった実験が可能になる。例えば話速を変化させる場合には原音声と雑音の時間比率ではなく、原音声が表示された継続時間の合計が本質的なパラメータとなる可能性も高い。また、イントネーションやリズムなどのパラ言語情報を有効に活用することで、人間の知覚に対してトップダウン的なヒントを多く与えることができ、一方で従来の音声認識ではこれらの情報は無視されやすいことから、能力のギャップ拡大に貢献できると期待される。

## 6. まとめと今後の課題

本報告では音声 CAPTCHA の設計手法について考察し、予備実験の結果について述べた。

今後の課題として以下が挙げられる。

- 現在実験に用いている音声について、マスキングに用いたホワイトノイズの音量が大きすぎて利用者が苦痛ではないか、という（視覚障害者からの）指摘があった。そこで、適切なマスキング方法について再検討を行う。

- 原音声の割合に関して 30%, 50%, 70% という条件をさらに細かくした実験を行う。

- 「混合法」と「削除法」の使いやすさを比較するために、タスク達成率、所要時間、心的負荷などの比較実験を行う。

- 音声に含まれる数字の桁数を教示する場合としない場合で了解度や認識率はどのように変わるかを検討する。

- 数字ではなく文字列や文章などでも適切な課題を実現できるかを検討する。

- 実際の利用状況に近い音声 CAPTCHA システムを試作し、特に高齢者や視覚障害者による評価を行う。

- 音韻修復に関する先行研究の知見を取り入れる、混合法と削除法の併用を試みる、など、手法の改良を行う。

さらに本研究は以下のような課題に発展すると考えられる。

- 音声認識されにくく人間にはちゃんと聞き取れる音声とはどのような音声なのか？

- 音声認識されにくい声を合成・加工する信号処理とはいかなるものであるか？

- どの程度「破る立場」からの検討を行えば十分に破られにくいと主張できるのか？

- 本報告では論じなかった「混合法」について、複数話者発話、多言語音声、残響、雑音、といった要素について設計手法を体系的に論じることや、難易度をパラメトリックに制御することが可能であるか？

- 人間には雑音にしか聞こえないが音声認識システムには言語音声として認識されるような音を作ることは可能か？

- 人間が言語音声を聴取する場合にトップダウン情報とボトムアップ情報がどのように用いられるのか？

- 統計的手法による音声認識技術の「裏をかく」ためにはどのような発想が必要なのか？

また CAPTCHA システムを別の問題解決に使う、という興味深い提案もある。例えば Carnegie Mellon University のグループが提案する reCAPTCHA システム<sup>(注10)</sup>は、実在する紙の本をスキャンして OCR で読み取ろうとした単語のうち認識に失敗した画像を提示する。さらに既知の課題と未知の課題を合わせて提示する。これにより CAPTCHA を「人力 OCR」として利用することを提案している。音声 CAPTCHA においてもこのアイディアは応用可能であろう。

謝辞

渡辺哲也氏（独立行政法人 国立特別支援教育総合研究所）からは本研究開始にあたって貴重な助言を得た。また本研究の一部は種村諒氏（東京大学工学部計数工学科）、2008 年度全学体験ゼミ「音楽・音響の信号処理と情報処理」受講生の吉里幸太氏・平川淳氏・岡場翔一氏の協力を得た。また小野順貴講師と嵯峨山・小野研究室メンバー各位、UAI 研究会および音声・音楽研究会（音音研）の参加者各位など多くの方々との議論から重要な示唆を得た。

## 文 献

- [1] Jim Thatcher, Michael R. Burks, Christian Heilemann, Shawn Lawton Henry, Andrew Kirkpatrick, Patrick H. Lauke, Bruce Lawson, Bob Regan, Richard Rutter, Mark Urban, Cynthia D. Waddell 著, UAI 研究会 翻訳プロジェクト 訳, 渡辺 隆行, 梅垣 正宏, 植木 真 監修: Web アクセシビリティ - 標準準拠でアクセシブルなサイトを構築 / 管理するための考え方と実践 -, 毎日コミュニケーションズ (2007).
- [2] Jonathan Holman, Jonathan Lazar, Jinjuan Heidi Feng, John D'Arcy: "Developing usable CAPTCHAs for blind users," Proceedings of the 9th international ACM SIGACCESS conference on Computers and Accessibility, Tempe, Arizona, USA, pp. 245 - 246, 2007.
- [3] 柏野牧夫: "音韻修復 - 消えた音声を修復する脳 -, " 日本音響学会誌 61 巻 5 号, pp.263-268, 2005.
- [4] R.M.Warren: *Auditory Perception: A New Analysis and Synthesis*, Cambridge University Press, Cambridge, 1999.
- [5] G.A. Miller, J.C.R. Licklider, "The intelligibility of interrupted speech," J. Acoust. Soc. Am., 22, 167-173 1950.
- [6] M. Nakayama, et al.: "CENSREC-4: Development of Evaluation Framework for Distant-talking Speech Recognition under Reverberant Environments," Proc. Interspeech, Sep 2008.
- [7] S. G. Hart, L. E. Staveland: "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research," in P. A. Hancock and N. Meshkati (Eds.) *Human Mental Workload*, Amsterdam, North Holland Press (1998).
- [8] 芳賀 繁: *メンタルワークロードの理論と測定*, 日本出版サービス, 2001.

(注10): reCAPTCHA, キャプチャを利用した人力高性能 OCR,  
<http://labs.cybozu.co.jp/blog/akky/archives/2007/05/recaptcha-human-group-ocr.html>  
<http://recaptcha.net/>