音韻修復効果を用いた音声CAPTCHAの検討

福岡　千尋[*1]　西本　卓也[*2]　渡辺　隆行[*1]

Investigation of speech CAPTCHA system using phonemic restoration

Chihiro Fukuoka[*1]　Takuya Nishimoto[*2] and Takayuki Watanabe[*1]

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. 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.

Keywords : Internet, Security, Visual Imparement, Speech Recognition, Speech Synthesis

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[1]．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[2]．

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[3]．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[4]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．

1. 音声CAPTCHAの設計方針

1.1 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

*1: 東京女子大学 現代文化学部 コミュニケーション学科
*2: 東京大学大学院 情報理工学系研究科
*1: Department of Communication, Tokyo Woman's Christian University
*2: Graduate School of Information Science and Technology, The University of Tokyo

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/

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

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

4: http://spamfizzle.com/CAPTCHA.aspx

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．

1.2 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 CHAPTCHA を簡単な手法で破ったとするブログ記事 [5] 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．

1.3 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 [6] 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．

1.4 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．

5: http://blog.wintercore.com/?p=11
6: http://research.microsoft.com/asirra/

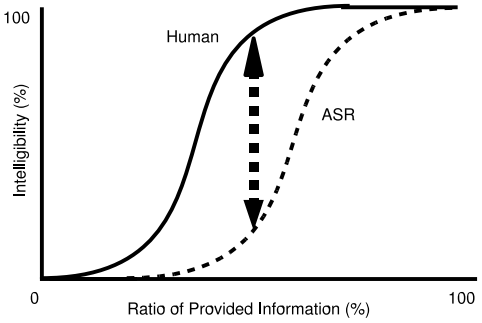

図 1　音声 CAPTCHA 設計の概念モデル

1.5　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けることができるかどうかが，削除法の有効性の判断基準となる．

2.　実験方法

2.1　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．

2.2　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．

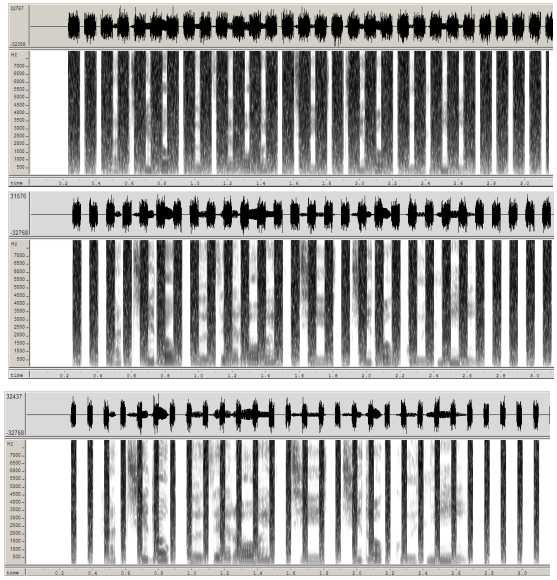
図 2　使用する音声ファイルの例．7 桁の数字読み上げ音声に対して，上から順に 30%,50%,70%の割合で原音声を残し，ホワイトノイズを埋めたもの．

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こりにくくなっている．

2.3　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 [7] を用いた．3 条件のそれぞ

7: http://htk.eng.cam.ac.uk/

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．

2.4 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．

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．

2.5 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．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．具体的にはリハーサルおよび複数の課題における各項目の評価の画面において過去の評定値を参照できるようにした．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．

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%

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

2.6 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．

最初に音量の確認を兼ねたリ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定を行わせた．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 つの尺度の負荷評定を行わせ，(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．

3. 実験結果と考察

3.1 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．

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

3.2　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．

3.3　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．

図 3 に実験結果のまとめを示す．

4.　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．

現在，文章の聴き取り課題においても削除法による音声 CAPTCHA が可能であるか，といった検討を行っている．具体的には，アナウンサーが文章を読み上げた音声を削除法で加工して，被験者に書き取らせる実験を行い，単語了解度による評価を行っている．その際に，読み上げ音声の話速の影響についても検討を行っている．また「聴き取りにくさ」[9] の主観評価によって，心的負荷評価と同じような効果が得られるのではないかと期待している．

さらに録音音声の代わりに，音声対話技術コンソーシアム (ISTC) [8] のもとで開発されている HMM 音声合成エンジン GalateaTalk [9] の合成音声を用いることも検討している．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．

将来は以下の課題に取り組む予定である．

(1) 混合法および削除法の音声 CAPTCHA 課題の難易度を音響信号から予測する手法を検討する．定常および非定常の雑音下音声了解度予測モデル [10], [11] などを参考に，音声が雑音にマスキングされている割合をサブバンドごとに計算し，重み付き平均を求める，といった手法をベースに検討する．

(2) 音声 CAPTCHA を聴取する人間の聴覚は，限られた音響刺激から効率よく心的辞書を検索し，単語や文章を聞き取っている．このプロセスを理解し適切に利用することは音声 CAPTCHA 設計において重要である．特に近年は TRACE や Shortlist といった心的辞書アクセスのモデルについて，アイトラッキングなど新しい技術が導入され興味深い研究がなされている [12]．また，シラブルを単位とする欧米言語と異なり，日本語の聴取においてはモーラの知覚が重要な役割を果たしている [13]．これらの知見を踏まえて音声 CAPTCHA 課題における認知プロセスの解明を目指す．

(3) Carnegie Mellon University のグループが提案する reCAPTCHA システム [10] 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呈示する．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．

reCAPTCHA の音声版は，機械では認識できない音声のディクテーションを支援する Web 2.0 的システムの実現可能性につながる．ウェブで公開されているポッドキャスト番組の音声認識を行い，利用者に誤認識の訂正を促す提案として PodCastle システムがある (http://podcastle.jp)．しかし，ユーザに誤認識訂正の動機付けを与えることは容易ではない．reCAPTCHA のアイディアを用いることで，ウェブを利用するために音声を書き起こす，という動機付けを与えることができる．

具体的には CMU の画像版 reCAPTCHA を踏襲し，ユーザに 2 つの音声を呈示する．片方の音声はすでに正解がわかっている音声である．もう 1 つの音声はシステムがまだ正解を知らない音声である．ユーザはこれらの 2 つの音声を聞き取って，テキストを入力する．前者の音声について正しいテキストを回答していたならば，ユーザが人間であると判断し，もう 1 つの

8: http://www.astem.or.jp/istc/
9: http://sourceforge.jp/projects/galateatalk/
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/

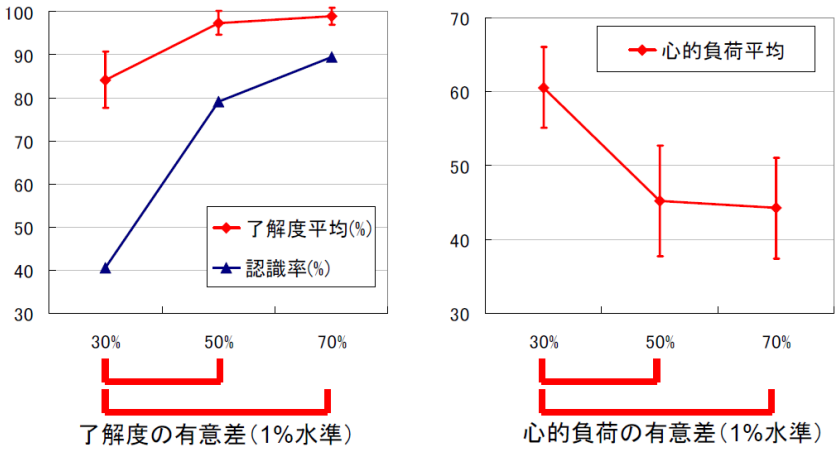

図 3　認識率・了解度および心的負荷の平均と標準偏差．横軸は原音声の割合．

音声についてもユーザの入力を新たに正解として記憶する．前者の音声が正しく回答されなかった場合には，入力者は人間ではなくソフトウェアロボットであると判断し，後者の入力を破棄する．実際には，未知の音声について（人間だと判断された）複数のユーザからの情報を統合し，信頼性の高い結果を得る必要があるだろう．

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西亀健太氏（東京大学大学院情報理工学系研究科），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・嵯峨山茂樹 教授・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．

参考文献

[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 interupted 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.

[9] 佐藤逸人, 森本政之, 佐藤 洋: "" 聞き取りにくさ "による音声伝達性能の評価," 日本音響学会誌 63 巻 5 号, pp.275-360, May 2007.

[10] ANSI/ASA S3.5-1997 "American National Standard Methods for Calculation of the Speech Intelligibility Index," American National Standards of the Acoustical Society of America (R2007) .

[11] K.S. Rhebergen, N.J. Versfeld, "A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," J. Acoust. Soc. Am. 117 (4), Pt. 1, April 2005.

[12] Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., Sedivy, J.E.: "Integration of visual and linguistic information in spoken language comprehension," Science, 268, 1632-1634 (1995).

[13] Cutler, A. and Otake, T. : "Rhythmic categories in spoken-word recognition," Journal of Memory and Language, 46-2, 296-322 (2002).