

# 音声・言語素性を用いた機械学習による 認知症進行度推定に関する予備的検討

外山 翔平<sup>1,a)</sup> 齋藤 大輔<sup>1</sup> 峯松 信明<sup>1,b)</sup> 小林 朗子<sup>2</sup> 小島 敏昭<sup>2</sup>

**概要:** 急速な社会の高齢化に伴い、認知症者数も増加の一途を辿っている。認知症の決定的な治療法が見つかっていない現状を鑑みるに、その早期発見に対する期待は大きい。本稿では口頭諮問による認知症スクリーニングである HDS-R (改訂長谷川式簡易知能評価スケール)、MMSE (Mini-Mental State Examination) などに対する回答や、諮問時の音声資料から抽出される音声・言語素性を機械学習に適用し、認知症の進行度である CDR (Clinical Dementia Rating) を推定する予備的検討を行った。

## A preliminary study of estimating clinical dementia rating based on machine learning methods using speech and language features

TOYAMA SHOHEI<sup>1,a)</sup> SAITO DAISUKE<sup>1</sup> MINEMATSU NOBUAKI<sup>1,b)</sup> KOBAYASHI AKIKO<sup>2</sup>  
KOJIMA TOSHIKI<sup>2</sup>

### 1. はじめに

我が国の急速な高齢化に伴い、様々な新しい社会サービスの必要性が叫ばれている。その一つが認知症者の介護である。認知症は症状進行によりコミュニケーション障害を生じ、本人の QOL のみならず、家族など、介護者の負担を大幅に増加させることもあり、社会全体とした取り組みが必要である [1]。また、認知症は決定的な治療法が確立されておらず、認知症の早期発見法の開発に対する期待は大きい。そのための取り組みとしては、脳活動など生体信号に基づくもの [2], [3]、当事者のビヘービア観測に基づくもの [4] などがあり、また、より簡便な早期発見法の可能性として、音声バイオマーカーを検討する研究も行われている [5], [6], [7], [8], [9]。更には、認知症と診断された方々を対象とした、韻律知覚に関する分析的検討も行われている [10]。上記したように認知症はコミュニケーション障害

を生じることもあり、介護者が (音声の韻律を通して) 伝えたいつもりでも当事者には伝わっていない、理解されていないことがあるからである。

さて、代表的な認知症スクリーニング手法として、改訂長谷川式簡易知能評価スケール [11] (以下、HDS-R と略す) や Mini-Mental State Examination [12] (同様に MMSE と略す) がある。これらはともに音声諮問を主とする診察であり、これらの諮問や脳画像の検診を通して総合的に専門医が診断を下す。認知症の進行度を示す度合いとしては、Clinical Dementia Rating (CDR) [13] スコアとして記述される。しかしながら、これらのスクリーニング手法に対して、1) 認知症の重症度が軽度あるいは初期段階の認知症例に対する感度が低い [14], [15], [16]、2) スコアが年齢や教育年数に影響を受ける [16]、3) スクリーニング実施に要する時間が長い [17]、4) 十分なトレーニングをうけた検査者でなければ実施が難しい [18]、5) 採点に際して主観的評価が入ることがある [18]、などの問題点が指摘されている。今後、高齢者を抱える家庭がますます増えることを考えると、信頼性の高い (予備的な) スクリーニングが家庭内で実施できることが望ましい。また、検査者が行うスクリーニングが音声による口頭諮問を主とすることを考える

<sup>1</sup> 東京大学大学院工学系研究科  
Graduate School of Engineering, The University of Tokyo  
Bunkyo-ku, Tokyo 113-8656, Japan

<sup>2</sup> 株式会社イフコム  
IFCOM Inc.

a) toyama@gavo.t.u-tokyo.ac.jp

b) mine@gavo.t.u-tokyo.ac.jp

と、同様の諮問を家庭内で行い、その諮問に正しく回答できたのか（即ち回答結果の正誤）、及び、その時の対話の様子（当事者の回答音声）を分析することで、ある程度の精度の検査ができることが望ましい。

以上を踏まえ本稿では、認知症音声スクリーニングの実現に対する予備的検討として、a) 回答結果の正誤、b) 回答音声に対する発話スタイル分析より定義される音声素性、c) 回答音声に対する音声認識結果より定義される言語素性を用い、各種機械学習法を用いて専門医による診断結果（CDR スコア）を予測する（後述するように実際には3クラスの識別問題として扱った）実験を行った。なお、本来であれば先行研究を精査し、各種バイオマーカーに基づく音声素性を検討する必要があるが、ここでは予備検討として、音声からの感情認識や発話スタイル認識で広く使われる音声特徴量である、openSMILE[19]のemobaseと呼ばれる特徴量を素性として使用した。また言語素性は、音声認識結果に対して、諮問の内容を考慮した上で用意した。

## 2. 利用したコーパスとタスク設定

本研究の遂行に際して、第四・五著者から提供された音声コーパス、テキストコーパスを用いた。これは諮問時の被験者（高齢者）の音声資料と、各諮問に対する回答の正誤に関するコーパスである（音声資料の書き起こしはない）。認知症は段階的なものであり、認知症の前段階である軽度認知障害（Mild Cognitive Impairment; MCI）であれば、治療をすることで認知症になることを遅らせることや避けることができる。そのため、認知症の早期発見には、MCIの段階での検出が重要である。CDR スコアと認知症の程度との対応は、健康（CDR 0）、認知症の疑い（CDR 0.5）、軽度認知症（CDR 1）、中等度認知症（CDR 2）、重度認知症（CDR 3）と定義されている。本研究は認知症の早期発見を目標としているため、CDR スコアとして0, 0.5, 1の被験者（総数1,550）を主たる対象とし、後述の実験もこの3クラスの識別問題として扱った。

被験者のデータとしては、専門医による診断結果（CDR スコア）、各種諮問に対する回答（正誤の様子）、その回答音声の生データ（切り出しなどはされていない）が利用可能である。諮問は下記の3種類より構成されている。

- 改訂長谷川式簡易知能評価スケール（HDS-R）、9 質問
- ミニメンタルステート検査（MMSE）、10 質問
- 物語再生、11 質問

各検査は幾つかの質問により構成されている（全30問）。個々の質問は、更に細かい小問に分かれ、これら部分点の総合点が回答の正誤スコアとなる。例えば、MMSEの質問1は日時の見当識を問うものであり（今年は何年ですか？今の季節は何ですか？今日は何曜日ですか？今日は何月何日ですか？5つの小問からなる）、年、季節、曜日、月、日に各々1点が割り振られ、その総合点が質問1の正誤ス

コアとなる。HDS-R、MMSEには日時の見当識を聞く問題以外に、場所の見当識（今どこにいますか？）を聞く問題も用意されている。更には、簡単な引き算や、物を見せて名前を言わせたり、文章を読ませたり、あるカテゴリの名前（例えば野菜）を思いつく限り言わせたり、一度覚えさせた項目を時間を経てから想起させたりする質問などで構成されている。即ち主として短期記憶、長期記憶、ワーキングメモリの働きを検査する質問より構成されている。なお、物語再生は、一度聞かせた物語を、再度再生させる質問である。これらの質問は、諮問の遂行、及び、得られた回答の正誤判定は一般家庭でもできる簡易な内容ばかりである。問題は、諮問の遂行ではなく、得られた回答からCDR スコアを推定することにある。

今回提供された回答の正誤コーパス及び音声コーパスは1,500名以上のコーパスとなり、後者は録音された音声そのまま提供されており、当初、以下のようなタスク設定をした。

- 3種類の諮問（HDS-R、MMSE、物語再生）に対する回答結果に対して、CDR スコア識別に対して貢献度が高い質問を、貪欲探索により絞り込む。
- 絞り込んだ質問に対する回答音声を、生音声から整備する。これは人手を使って行う。
- 整備された音声コーパスから各種音声素性、言語素性を抽出し、また、回答の正誤も使って再度CDR スコア識別実験を行う。

即ち、機械学習を用いたCDR スコア識別としては、1) 諮問の回答の正誤のみを用いた識別、2) 諮問の回答+音声素性・言語素性を用いた識別、の2段階の実験を行った。以下の節で、各実験について詳細を説明するが、HDS-Rの質問 $n$ を $HQ_n$ 、MMSEの質問 $n$ を $MQ_n$ 、物語再生の質問 $n$ を $SQ_n$ と示す。また、複数の小問から構成される質問があるため、質問ごとに満点が異なるが、以下では、点数の正規化は行っていない。

## 3. 回答の正誤コーパスのみを用いたCDR スコア識別と質問選択

### 3.1 各質問とCDR スコアとの相関

まず初めに、個々の質問の回答の様子がCDR スコア（但し、0, 0.5, 1のみを採択）とどの程度連動しているのを見るために、両者の相関係数を計算した。なお、 $HQ$ 、 $MQ$ 、 $SQ$ 全ての諮問を受けた（かつ、本検討時利用可能であった）754人のデータを用いた。相関の高い5質問を表1に示す。 $MQ_1$ 、 $MQ_2$ 、 $HQ_2$ はいずれも日時や場所の見当識であり、見当識を問う質問がCDRの識別に有効であることが示唆される。また $HQ_7$ 、 $MQ_5$ はいずれも3単語の遅延再生であり、認知症患者はワーキングメモリの処理に難があるという知識と合致する。

表 1 相関の高い質問項目

| 質問  | 相関     |
|-----|--------|
| MQ1 | -0.760 |
| MQ2 | -0.728 |
| HQ2 | -0.728 |
| HQ7 | -0.686 |
| MQ5 | -0.663 |

表 2 全質問を用いた CDR 識別

| 識別器                       | 識別精度 |
|---------------------------|------|
| ロジスティック回帰                 | 0.78 |
| Support Vector Classifier | 0.77 |
| Random Forest             | 0.75 |
| Neural Network            | 0.74 |

### 3.2 全質問を用いた CDR スコアの識別実験

HQ, MQ, SQ の全質問の回答の様子を各被験者の入力素性として、CDR スコア (0, 0.5, 1) を識別するタスクを、機械学習の各種アルゴリズムを用いて遂行した。利用可能な全データよりランダムに 8 割を抽出して学習データとし、残りをテストデータとした。識別結果を表 2 に示す。結果より約 75% の精度で識別が可能であることが分かる。なお、MQ, SQ を受けた被験者に対して (MQ, SQ だけを使って) 同様の実験を行ったところ、同様の結果を得た。HQ, MQ の質問項目から容易に推測されることではあるが、両質問セットの間には高い冗長性がある。

### 3.3 CDR スコア識別に寄与する三質問の選定

どの質問に対する回答音声を整備すべきかを定めるため、利用できる質問の中から 3 つの質問を選定することとした。質問総数は 30 あるため、ここから任意の 3 つの質問を選ぶ組み合わせの数は約 2 万通りあるが、全組み合わせに対して、5 種類の識別器 (K-neighbors classifier, support vector classifier, Gaussian process classifier, Random forest classifier, AdaBoost classifier) を適用し、第 3.2 節と同様の実験を繰り返し、識別率の高い三質問と識別器の組み合わせを選定した。なお識別器は scikit-learn ライブラリ [20] を用いて準備した。表 3 に識別率が 67% (即ち 2/3 以上) となったものを示す。本実験では、細かなパラメータチューニングは行ってないためか、単純かつサンプルベースの K-neighbors classifier が安定した精度を示している。採択された質問としては、MQ2, HQ2, HQ9 が目立つ。更に、上位 100 位までの組み合わせに登場した各質問の登場回数を数えると、表 4 が得られた。特に、MQ2 (場所の見当識), HQ2 (日時の見当識) は約 70% のケースで採択されており、CDR 識別タスクにおいて、重要な質問となることが予想される。MQ1 (日時の見当識) や HQ7 (三単語の遅延再生) の出現回数も多いが、これは第 3.1 節での議論と合致する。

表 3 三質問による CDR 識別結果

| 三質問           | 識別器              | 識別率  |
|---------------|------------------|------|
| HQ9, MQ2, MQ8 | K-Neighbor       | 0.68 |
| HQ9, MQ2, MQ3 | K-Neighbor       | 0.68 |
| HQ2, MQ2, MQ4 | Gaussian Process | 0.68 |
| HQ2, MQ2, MQ4 | AdaBoost         | 0.68 |
| HQ2, HQ9, MQ1 | K-Neighbor       | 0.68 |
| HQ9, MQ2, SQ9 | K-Neighbor       | 0.67 |
| HQ9, MQ1, SQ7 | K-Neighbor       | 0.67 |
| HQ5, HQ9, MQ2 | K-Neighbor       | 0.67 |
| HQ1, HQ9, MQ2 | K-Neighbor       | 0.67 |

表 4 上位 100 位に高頻度で出現した質問

| 質問  | 出現回数 |
|-----|------|
| MQ2 | 70   |
| HQ2 | 69   |
| HQ9 | 26   |
| MQ1 | 19   |
| HQ7 | 9    |
| MQ7 | 9    |

登場回数が 9 回以上であった 6 質問間の相関係数も計算し、更には、各質問の質問意図を考慮し、音声資料を整備すべき三質問を選定した。まず、CDR との相関が高く、CDR 識別実験でも多くの場合において採択されていた HQ2 (日時の見当識), MQ2 (場所の見当識) を採択した。MQ1 は HQ2 との相関が 0.9 以上あるため省いた。残りの HQ9 (言語の流暢性), HQ7 (三単語の遅延再生), MQ7 (文の復唱) から残り一つを選ぶこととした。HQ9 は任意の野菜を思い出す限り答えさせる質問であり、音声技術を導入する場合、音声認識辞書を拡張する必要がある。HQ7 で使われる単語は 6 種類に限られており、また、MQ7 は一種類の文の復唱であり、音声技術の導入は比較的容易である。最終的に CDR スコアとの相関を考慮し、HQ7 を採択し、HQ2 (日時の見当識), MQ2 (場所の見当識), HQ7 (三単語の遅延再生) の回答音声を手動で整備することとした。なお、この三質問を用いた場合、Random Forest Classifier, Gaussian Process Classifier, AdaBoost Classifier を用いた場合に 64% の精度を示している。また、三種類の諮問の中から上記の三質問を選択したことを専門医に伝えたところ、彼らの知見とも合致するものであった。

## 4. 選定された質問とその音声・言語素性を用いた CDR スコア識別

選定された HQ2, MQ2, HQ7 の質問に限定し、音声素性、言語素性の利用も含め、CDR スコア識別実験を遂行する予定であったが、音声コーパスの整備が完了したのは HQ2 (日時の見当識) のみであったため、音声・言語素性の利用は、HQ2 のみで行った。また、第 3 節の実験では当初利用可能であったデータを 4:1 として学習及び評価に用い

表 5 使用した識別器

| 識別器                     | 略称  |
|-------------------------|-----|
| KNeighborsClassifier    | KN  |
| DecisionTreeClassifier  | DT  |
| LinearSVCClassifier     | LSV |
| SupportVectorClassifier | SV  |
| RandomForestClassifier  | RF  |
| MLPClassifier           | MLP |
| AdaBoostClassifier      | AB  |
| GaussianNaiveBayes      | GNB |

表 6 全質問 (30 問) を使った識別再実験

| 識別器 | 精度   | 識別器 | 精度   |
|-----|------|-----|------|
| KN  | 0.70 | RF  | 0.70 |
| DT  | 0.66 | MLP | 0.73 |
| LSV | 0.71 | AB  | 0.70 |
| SV  | 0.74 | GNB | 0.58 |

たが、本節では、利用できるデータ数が1,239名 (CDR0, 0.5, 1 に対して各々 362, 434, 443 名) に増え、また、10 分割交差検定を行い、結果の信頼性を高めることを努めた。また、識別器としては、scikit-learn ライブラリ [20] の表 5 の 8 つの識別器を使用した。また Grid Search によるハイパーパラメータの最適化も施した。

#### 4.1 全質問を用いた CDR スコアの識別再実験

データが拡張されたこともあり、HQ, MQ, SQ 全質問 (30 種類) の回答の正誤コーパスのみを用いた CDR スコア識別実験を再度行った。結果を表 6 に示す。第 3 節の実験結果 (表 2) と比べて、若干精度が落ちる結果となった。

#### 4.2 音声素性を用いた CDR スコアの識別実験

第 1 節で述べたように、認知症者の音声の韻律的特徴に特有のパターンが観測されるなどの報告もあり、ここでは、感情・発話スタイル認識などのタスクで広く利用される openSMILE ツールキットで使われている emobase 特徴量を、音声素性として利用した識別実験を行った。

openSMILE では、フレーム単位で定義される (抽出される) 局所的な音響特徴量に対して、より長い時間長を単位とした (例えば発話など) 統計量を計算し、最終的な音響特徴量セットを導出している。局所の特徴量としては、波形の RMS, F0, MFCC, 有声度, 更にはゼロ交差率などがある。これらに対して、最大値, 平均値, 標準偏差, レンジ (最大値-最小値), 更には最大値が観測された発話内の相対位置など、より長い時間長を有する発話を単位とした統計量を導出する。標準特徴量である emobase では 988 次元の特徴量を各発話ごとに抽出することになる。なお、対象とした回答音声は HQ2 (日時の見当識) のみであり、より詳細には、「今日は何年の何月何日ですか? 何曜日ですか?」という質問に対する回答となる。HQ2 では、

表 7 音声素性を用いた識別結果

| 識別器 | 精度   | 識別器 | 精度   |
|-----|------|-----|------|
| KN  | 0.51 | RF  | 0.54 |
| DT  | 0.50 | MLP | 0.54 |
| LSV | 0.51 | AB  | 0.55 |
| SV  | 0.47 | GNB | 0.55 |

表 8 HQ2 回答のみを用いた識別結果

| 識別器 | 精度   | 識別器 | 精度   |
|-----|------|-----|------|
| KN  | 0.61 | RF  | 0.63 |
| DT  | 0.63 | MLP | 0.63 |
| LSV | 0.63 | AB  | 0.63 |
| SV  | 0.63 | GNB | 0.56 |

年, 月, 日, 曜日各々の正誤に対して 1 点が割り当てられ、合計 4 点の質問である。

HQ2 の音声素性を用いた結果を表 7 に示す。なお、4 点満点の HQ2 スコアのみを用いた (音声素性を用いない) 識別結果についても表 8 に示す。この実験結果からは、emobase を単純に用いるよりは、1 次元の HQ2 のスコアの方が CDR の識別には有効であると言える。即ち、emobase をそのまま用いた音声素性では、認知症らしさを捉えるには不十分な素性定義であると言える。

#### 4.3 言語素性を用いた CDR スコアの識別実験

第 4.2 節では、HQ2 回答音声に「どのような発話スタイル」だったのか、に着眼した識別実験を行った。本節では、音声認識器を利用し、HQ2 回答音声に「どのような単語が現れたのか」に着眼して言語素性を定義し、これらを用いて識別実験を行った。

HQ2 は日時の見当識を問う質問であり、回答音声には西暦, 年号, 月, 日, 曜日などのキーワードが観測されることになる。なお、音声認識器は、CSJ コーパス [21] を用いて Kaldi ツールキット [22] により構築した。このコーパスは主として学会講演からなる日本語話し言葉コーパスである。話者の年齢は 20 代から 60 代の健常者であり、高齢者の音声に適した認識器とはなっていない。以下に認識結果例を CDR スコア別に示す。なお、HQ2 回答音声は書き起こされていないため、認識精度は算出してない。

CDR0: あ えー <sp> えーつとー <sp> 平成で言うと二十三年の<sp> 十月うーうす  
 CDR0: <sp> へい 千二十四年の一月三十日ですよ  
 CDR0.5: 平成二十三年<sp> 自由に血のず<sp> えつと<sp> の<sp> ふ 使いですか  
 CDR0.5: きゅー約あの一何面とうえーと昭和じゃなくて<sp> 平成<sp> 十五年<sp> えーになんつうの<sp> えーにゆに中の  
 CDR1: 今日の避難になる訳ないんで<sp> ま知れました

表 9 13 種類のキーワード

|                              |
|------------------------------|
| 年, 年, 日, 平成, 昭和, 二十          |
| えー, あの一, あつ, あー, うー, ん, (sp) |

表 10 言語素性を用いた識別結果

| 識別器 | 精度   | 識別器 | 精度   |
|-----|------|-----|------|
| KN  | 0.51 | RF  | 0.55 |
| DT  | 0.54 | MLP | 0.53 |
| LSV | 0.43 | AB  | 0.56 |
| SV  | 0.52 | GNB | 0.52 |

CDR1: そのですね <sp> A とあ <sp> んねんになつたものをおー <sp> 自由にあ数を <sp> あーはー <sp> 困りましたねえー <sp> 十一があつぬ競売ふんじ十日ですか <sp> と

全体的な傾向としては、CDR が低いほど正しく答えられており、認識結果も「読みやすく」なっている。CDR が高い被験者の場合、発話時間も長くなり、また、認識誤りも増加するように見受けられた。内容としても、当日の年月日を聞いているにも拘わらず、自身の生年月日を答えたり、また、フィルターも増加している様子が散見された。音声認識器の（音響的な）学習環境と、今回の被験者音声とのミスマッチは避けられないため、全体的な認識率は低くならざるを得ないが、下記実験は、上記の認識結果に対して各種言語素性を定義して行った。

用いた言語素性としては、2 種類からなる 17 次元の素性を用意した。まず認識結果より得られる対数尤度、発話時間、発話単語数、発話頻度である。また、幾つかの発話を聴取し、表 9 のキーワードを定義し、これらが「認識結果中に一回以上観測されたか否か」を示す素性を用意した。

結果を表 10 に示す。音声素性のみを用いた結果（表 7）と類似した性能が示されており、また、両素性に対して良好な精度を出しているのは RF や AB であるが、1 次元である HQ2 スコアのみを用いた場合の精度よりも低い値となった。認識精度の低さは否めず、CDR スコアを識別するための素性としては不十分なものであることが分かる。

## 5. 音声・言語素性及び回答コーパスを用いた CDR スコアの識別実験

これまで検討してきた、回答コーパスの各質問（HQ, MQ, SQ）の回答の正誤、及び、HQ2 の回答音声から抽出される音声素性、更には、HQ2 回答音声の音声認識結果から定義される言語素性を用いた CDR スコアの識別実験を行った。全質問 30 問に対する回答に対して HQ2 の音声・言語素性を追加した場合の結果を表 11、及び、HQ2 のみの回答に対して HQ2 の音声・言語素性を追加した場合の結果を表 12 に示す。なお、全 30 問の回答のみを用いた識別性能は表 6、HQ2 の回答のみを用いた識別性能は表 8

表 11 全質問+音声・言語素性による識別結果

| 識別器 | 精度   | 識別器 | 精度   |
|-----|------|-----|------|
| KN  | 0.51 | RF  | 0.59 |
| DT  | 0.63 | MLP | 0.54 |
| LSV | 0.53 | AB  | 0.69 |
| SV  | 0.34 | GNB | 0.59 |

表 12 HQ2 + 音声・言語素性による識別結果

| 識別器 | 精度   | 識別器 | 精度   |
|-----|------|-----|------|
| KN  | 0.51 | RF  | 0.59 |
| DT  | 0.62 | MLP | 0.54 |
| LSV | 0.51 | AB  | 0.64 |
| SV  | 0.34 | GNB | 0.58 |

であり、これらが比較対象となる。全質問の回答を用いた場合、DT, AB, GNB は音声・言語素性を追加しても精度の変化は小さかったが、他の識別器では精度を劣化させる結果となった。HQ2 の回答のみを用いた場合も、DT, AB, GNB は精度の変化は小さいが、他の識別器では精度は劣化した。なお、AB, GNB の場合は、僅かであるが精度は向上した。今回の実験では、諮問回答に基づく素性の次元数は 30 あるいは 1 であり、それに対して、約 1000 次元の音声・言語素性を追加している。次元数のアンバランスも結果に影響していると思われる。

## 6. まとめと今後の課題

本研究では、認知症スクリーニングで広く使われている、改訂長谷川式簡易知能評価スケール、ミニメンタルステート検査などに対して、これらが実際に適用された際の各質問に対する回答の正誤データ、更には、被験者の回答音声を用いた、CDR スコアの自動識別を検討した。本研究は認知症の音声スクリーニングを目的とした予備的検討として位置づけており、回答音声から抽出した素性としては非常に簡便なもの（ナイーブなもの）を使用した。音声素性としては openSMILE の emobase、言語素性としては CSJ-Kaldi による認識結果に対するキーワード素性を用いた。また識別器としては scikit-learn ライブラリが提供する各種識別器を用いた。なお、回答音声コーパスの整備の都合上、音声・言語素性は、HQ2 の音声のみを用いた。結果として、全 30 問を用いた場合の精度は 0.74、HQ2 だけを用いた場合の精度は 0.63 であった。後者に対して HQ2 の音声素性、言語素性を加味した場合の精度は 0.64 となり、精度向上は極めて僅かであった。

今後の課題としては、HQ2 以外の音声コーパスの整備、先行研究において示されている高齢者、特に認知症であることによる特有な音声バイオマーカーを積極的に利用した音声素性の導入、S-JNAS など高齢者音声コーパスを用いた適応処理による、より高精度な認識器の導入とそれに伴う、言語素性の高洗練化が上げられる。今回識別対象とし

た CDR であるが、専門医は、各種諮問の結果、被験者の音声の聴取、更には、被験者の音声発声時のビヘービア観測、脳画像に基づく検診、家族の方々から提供される情報を参照しつつ、総合的に診断している。機械学習と専門医とを公平に比較するためには、諮問に対する回答と音声聴取のみによる診断を行ってもらう必要もあるだろう。また、認知症の簡易スクリーニングシステムが本研究の最終目的であるが、この場合のスクリーニングの現場は、病院ではなく自宅となる場合が多い。各種の諮問を検査者が病院で行う場合と、家族の者が自宅で行う場合とでは、被験者の回答、発声の様子も異なってくると思われる。このような差異をどう扱うのかについても、検討する必要がある。

## 謝辞

各種実験の遂行に際し、峯松・齋藤研究室の機械学習勉強会メンバーの協力を得た。また、最終的な実験結果の整備に際し、小杉康宏氏の協力を得た。ここに感謝する。

## 参考文献

- [1] 厚生労働省, “認知症の医療と生活の質を高める緊急プロジェクト”, 成果報告書 (2008)  
<http://www.mhlw.go.jp/houdou/2008/07/h0710-1.html>
- [2] 上田他, “脳波解析によるアルツハイマー病の特性に関する研究”, 電気学会論文誌 C, vol.130, no.10, pp.1827–1832 (2010)
- [3] 加藤他, “課題実行時 fNIRS 脳機能計測データのベイジアンマイニングに基づく認知機能障害の 3 群判別”, 人工知能学会論文誌, vol.27, no.2, pp.28–33 (2012)
- [4] 阿部他, “行動センシングによる認知症の早期発見システム”, 第 13 回情報科学技術フォーラム論文集 (FIT2014), pp.299–300 (2014)
- [5] 加藤他, “高齢者音声韻律特徴を用いた HDS-R スコアとの相関分析”, 人工知能学会論文誌, vol.26, no.2, pp.347–352 (2011)
- [6] V. Taler *et al.*, “Comprehension of grammatical and emotional prosody is impaired in Alzheimer’s disease,” *Neuropsychology*, vol.22, no.2, pp.188–195 (2008)
- [7] K. Hoyte *et al.*, “Components of speech prosody and their use in detection of syntactic structure by older adults,” *Experimental Aging Research*, vol.35, no.1, pp.129–151 (2009)
- [8] Aharon *et al.*, “Speech-based automatic and robust detection of very early dementia,” *Proc. INTERSPEECH*, pp.2538–2542 (2014)
- [9] 黒川他, “音響的特徴を利用した高齢者の認知症傾向の分析”, 日本音響学会春季講演論文集, 1-Q-36, pp.313–314 (2017)
- [10] 齋藤他, “感情韻律が認知症高齢者の音声認知に与える影響”, 日本音響学会春季講演論文集, 1-10-13, pp.1161–1164 (2017)
- [11] 加藤他, “改訂長谷川式簡易知能評価スケール (HDS-R) の作成”, 老年精神医学雑誌, vol.2, no.11, pp.1139–1347 (1991)
- [12] M. Folstein *et al.*, “Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician,” *Journal of Psychiatric Research*, vol.12, no.3, pp.189–198 (1975)
- [13] J. C. Morris *et al.*, “The Clinical Dementia Rating (CDR): Current version and scoring rules,” *Neurology*, vol.43, no.11, pp.2412–2414 (1993)
- [14] M. Ijuin *et al.*, “Validation of the 7-minute screen for the detection of early-stage Alzheimer’s disease,” *Dementia Dement Geriatr Cognitive Disorder*, 25, 248–255 (2008)
- [15] 伊集院他, “軽度アルツハイマー型認知症例に対する MIS (Memory Impairment Screen) の適用可能性”, 老年精神医学雑誌, 19, pp.1349–1356 (2008)
- [16] T. Tombaugh *et al.*, “The minimal state examination: a comprehensive review,” *J. Am. Geriatr Soc.*, 40, pp.922–935 (1992)
- [17] H. Brodaty *et al.*, “What is the best dementia screening instrument for general practitioners to use?” *Am. J. Geriatr Psychiatry*, 14, pp.391–400 (2006)
- [18] 伊集院陸雄, “軽度アルツハイマー型認知症に対するスクリーニング・ツール”, 人工知能学会全国大会論文誌, 2F1-NFC4-11 (2009)
- [19] openSMILE:  
<http://audeering.com/technology/opensmile/>
- [20] scikit-learn:  
<http://scikit-learn.org/>
- [21] CSJ:  
[http://pj.ninjal.ac.jp/corpus\\_center/csj/](http://pj.ninjal.ac.jp/corpus_center/csj/)
- [22] KALDI:  
<http://kaldi-asr.org>