

別添3)

厚生労働科学研究費補助金

(政策科学総合研究事業(臨床研究等 ICT 基盤構築・人工知能実装研究事業))

総括研究報告書

電子カルテと連携する音声認識システムのニーズ把握及び
音声認識システムに用いられる医療用語辞書の編纂に関する研究

研究代表者 野田 和敬 千葉大学医学部附属病院 総合診療科 助教

研究要旨

診療業務の効率化は限られた医療資源への負荷を軽減する上で重要であり、それを達成するひとつの手段として、電子カルテと連携する音声認識技術の活用に期待が寄せられている。本研究では、アンケート調査により音声認識システムに対するニーズの高い領域・診療場面を明らかにし、次いで、ニーズの高い診療場面での辞書構築を図ることを目的とした。申請者らが平成 29・30 年度に医療現場で求められる音声認識システムのニーズについて調査したところ、主に次の 3 用途においてニーズが高かった。① 音声による電子カルテ操作(音声コマンドと呼ばれるもの)、② 救急対応時の処置等の記録や身体所見の記録などのハンズフリーでの用途、③ 医療面接や病状説明などの音声自動テキスト化、である。

そこで、ニーズが特に高く、かつ、実用化が期待できる音声コマンド(①)の実装を最優先課題と考え、「電子カルテ項目名辞書」の整備の一環として「電子カルテ項目名」の標準呼称を作成するとともに、「薬剤名辞書」の整備を行った。薬剤名辞書は医薬品データベースを元に医薬品の呼称のリスト化および、音声認識用辞書編纂を行った。編纂した薬剤名辞書の有効性を評価するために、音声認識精度における薬剤名辞書の効果を標準辞書と比較して検証したところ、口元マイクを用いた場合で約 36%、マイクロフォンアレーを用いた場合で約 32%の正答率の向上が得られた。

また、医療面接の音声自動テキスト化(③)の精度向上に今後必要となる要素についての分析を行った。医療面接の自動テキスト化において、音響モデルでは「収録音声データのみ」から構築したモデルを用いた場合に、言語モデルでは「日本語話し言葉コーパス(CSJ) + 医療面接時の書き起こしテキスト + カルテ記載テキスト」から構築したモデルを用いた場合に、最も認識精度が高かった。一般にモデル構築に用いるデータ量が多いほど高い認識率を示すと予想されるが、今回の検証では、約 240 時間分の CSJ データよりも約 40 時間分の収録音声データから構築した音響モデルの方が高い認識率を示したことは注目すべき点である。また、言語モデルでは「CSJ + 医療面接時の書き起こしテキスト」を汎用モデルとし、そこに特定領域のテキストデータ(今回は総合診療科カルテ記載

テキスト)を追加することによって連続音声認識をその領域に適応させられる、ということが示された。言語モデル構築に用いた元データ間の語数の差を補正する処理を追加することで、書き起こしテキストの効果がより強化される可能性があり、今後の課題である。

医師と患者の発話では、患者の発話の認識精度が低くなる傾向があった。収録機材による差として、マイクロフォンアレーよりもピンマイクを用いた場合の認識精度が高く、今回の評価対象ではビームフォーミング技術よりも発話源にマイクをできるだけ近づけることの方がより有利であった。また、実際の医療面接を収録した音声よりも、その書き起こしテキストの読み上げを収録した音声の場合で大幅な認識率の向上が得られたことから、発話の際の意識の違いも認識精度に大きな影響を与えることが示された。明瞭な発話については患者側にこれを負担させることは困難であることが予想されるが、医師側においては意識的に明瞭な発話を心がけることで実用的な音声認識精度が得られる可能性がある。医療面接時の音声自動テキスト化の精度向上に関する今後の対策として、1. より一層の言語モデル・音響モデルの改良、2. 歪みの少ないノイズ低減処理、3. 室内残響の低減、4. Voice Activity Detection(VAD)の高度化、が挙げられる。

②の救急場面以外でのハンズフリー用途としては、身体所見の記録というニーズが高かったことから、実地医家向けの身体所見用辞書の編纂を行った。一般財団法人医療情報システム開発センター(MEDIS-DC)のホームページにおいて公開されている「症状所見マスター<身体所見編>(2014.03.06 更新)」を用いて、実地医家で頻用されると思われる用語を選定し、音声認識を目的とした辞書編纂を行い、その有効性を検証した。身体所見用辞書の効果を標準辞書と比較して検証したところ、身体所見用語のみの認識では、口元マイクを用いた場合で約35%、マイクロフォンアレーを用いた場合で約36%の正答率の向上が得られた。所見用語と判定用語を組み合わせた身体所見パターンの認識もほぼ同等の精度が得られた。孤立単語認識を採用した場合、収録された用語であれば実用的なレベルで認識されることが実証できたことから、身体所見記載に用いられる用語を拡充すること、それらに付随する判定用語も合わせて充実させることが有益であると考えられる。しかし、実運用の場面では孤立単語認識だけでは対応できない可能性もあり、連続音声認識との併用などの対策が必要である。

研究分担者	所属研究機関	職名
生坂政臣	千葉大学医学部 附属病院 総合 診療科	教授
傳康晴	千葉大学大学院 人文科学研究院	教授
鈴木隆弘	千葉大学医学部 附属病院 企画 情報部	准教授
大平善之	国際医療福祉大 学 医学部	主任 教授
上原孝紀	千葉大学医学部 附属病院 総合 診療科	講師
島井健一郎	千葉大学医学部 附属病院 企画 情報部	特任 講師
中田孝明	千葉大学大学院 医学研究院	教授
新津富央	千葉大学大学院 医学研究院	講師

A. 研究目的

申請者らが平成 29・30 年度に医療現場で求められる音声認識システムのニーズについて調査したところ、主に次の 3 用途においてニーズが高かった¹⁾。① 音声による電子カルテ操作(音声コマンドと呼ばれるもの)、② 救急対応時の処置等の記録や身体所見の記録などハンズフリーでの用途、③ 医療面接や病状

説明などの音声自動テキスト化、である。

そこで平成 30 年度には、ニーズが特に高く、かつ、実用化が期待できる音声コマンド(①)の実装を最優先課題と考え、「電子カルテ項目名辞書」の整備に注力した。平成 31 年度はさらに「電子カルテ項目名」の標準呼称案を作成するとともに、「薬剤名辞書」の整備を行うこととした。

また一般に音声認識では、コンピューター相手の monologue (=ディクテーション、音声コマンド)よりも人相手の dialogue (=対話)において精度が劣る傾向にある。したがって、医療面接などの音声自動テキスト化(③)については、音源からマイクまでの距離や話し方の問題で入力音質が劣るために、現時点では実用的な精度を出すまでには至っていない。認識精度の向上には音声の入力環境に対する信号処理を行い、言語モデル/音響モデルおよび用語辞書を整備するなどの集学的なアプローチが必須である。しかし、③は幅広い領域でニーズが高いだけでなく、構造化カルテの作成という観点でも重要な技術となる可能性がある。

医師が作成する診療録は自由記載で記録されるのが一般的であり、二次利用しやすい構造化された記録を残すことが長年の課題となっているが、入力時に構造化記録を行うことはユーザーインター

フェースの問題もあってこれまでに成功例がない。既存の自由記載を自然言語処理で利活用しようとする試みも行われているものの、この方法では自由記載の中に有益な情報が記録されていなければそもそも対処できない。音声認識による自動テキスト化は、構造化カルテの作成支援においてこれらの問題点を解決する技術的基盤となる可能性がある。そこで平成 31(令和元)年度は、平成 29・30 年度の成果を踏まえて、医療面接の音声自動テキスト化の精度向上に今後必要となる要素を明らかにすることとした。

なお、②の救急場面での高いニーズについては、②以外の固有のニーズとして救急室での薬剤投与や初期検査のオーダーリング用途や、プレホスピタルでの情報伝達内容に関するテキスト書き起こしなどの要望があったが、前者は①、後者は③の延長線上にあるものと考えられたため、まずは①および③に注力することが必要と考えた。

さらに、②の救急場面以外でのハンズフリー用途としては、身体所見の記録というニーズが高かったことから、身体所見用の辞書編纂を実施することとし、まずは実地医家向けの身体所見用辞書編纂を行うこととした。

B. 研究方法

<音声コマンドに関する辞書編纂>

1) 電子カルテ項目名の標準呼称案の作

成

主要電子カルテベンダー〔病院向けの電子カルテシェアで 10%以上の病院数を占める富士通、ソフトウェア・サービス、シーエスアイ、日本電気、ならびに、音声認識技術の医療応用により業務の高度情報化の効果が期待される国立大学病院において稼働している電子カルテベンダー(日本アイ・ビー・エム、キヤノンメディカルシステムズ)〕から電子カルテ内に存在する機能および項目の名称を収集し、音声コマンドの普及促進のため電子カルテ機能の標準呼称案を作成した。

2) 薬剤名辞書

(1) 辞書編纂

医薬品データベースを元に医薬品の呼称のリスト化および、音声認識用辞書編纂を行った。医薬品データベースとしては当院で作成した医薬品マスターを用いた。医薬品マスターは定期的に更新されるため、マスターから音声認識用辞書編纂を作成する具体的かつ効率的な手法について検討し、構築した。

具体的な辞書編纂手法は以下のとおり。

- ① 医薬品マスターの「漢字名」から薬剤名と無関係と思われる部分(冒頭の「治)」や末尾の「50mg」などを削除した。
- ② あらかじめ定義した、「経口用」「筋注」などの用途を表す語や、「ツムラ」「三和」などの会社名などを、①の結

果の先頭から削除した。

- ③ あらかじめ定義した、「OD 錠」「カプセル」などの形状を表す語や、「点眼」「静注」などの処方を表す語などを、②の結果の末尾から削除した。
- ④ ③の結果を形態素解析(解析エンジン:MeCab + 解析辞書:UniDic)し、発音形を得た。
- ⑤ 医薬品マスターの「カナ」との照合から、④の結果の誤りを手作業で修正した。
- ⑥ ⑤の結果を「漢字名」「発音形」「オーダーコード」の表にまとめた。
- ⑦ ⑥の表を音声認識システム Julius の孤立単語認識用単語リストに変換した。

(2) 辞書の評価

編纂した薬剤名辞書の有効性を評価するために、音声認識精度における薬剤名辞書の効果を検証した。検証方法は、千葉大学医学部附属病院(以下、当院)での処方実績に基づいて高頻度に処方されている薬剤 200 件を選定し、3名の医師がそれぞれ読み上げ、その認識精度を測定した。音声入力には 2 台の Windows 端末に、i) 口元にセットした単一指向性マイク(サンワサプライ MM-MCU02BK)および ii) マイクロフォンアレー(ReSpeaker Mic Array v2.0)をそれぞれ接続して同時に行い、音声認識エンジンには Julius を用いた。なお、評価については Julius に付属している標準

辞書のみを用いた場合と標準辞書に薬剤名辞書を加えた場合とで比較し、音素列での一致度(ゲシュタルトパターンマッチング)によって認識精度の比較を行った。

<医療面接の音声自動テキスト化の精度向上に関する分析>

医療面接の音声自動テキスト化について、適用する辞書・言語モデルならびに音響モデルを変えて音声認識率を比較した。

1) データ収集

医療面接・病状説明時の自動テキスト化を達成するには、辞書編纂や言語モデル/音響モデルの構築が重要である。その構築に必要な実発話データとして、臓器横断的な診療を行っている総合診療科でのデータで代表させることとした。総合診療科での医療面接時の音声を収録し、辞書編纂やモデル構築に利用する元データとした。医師と患者の両者に対して、i)それぞれの胸元に取り付けたピンマイクによる収録と、ii)マイクロフォンアレーでのフロアマイクによる収録、とを同時に行った。使用した機材は 2)のとおりである。収録した音声は技術補佐員 2 名によりテキストへの書き起こし作業を行った。患者名などの個人情報はずべて伏せ字にした。これらの書き起こしテキストから辞書・言語モデルを構築するとともに、収録音声から機材 i)を用

いた場合と、機材 ii)を用いた場合のそれぞれで音響モデルを構築した。

さらに、学習データの追加による効果を検討するため、上記、医療面接場面の書き起こしテキストとは別に、カルテ記載テキストとして、平成 28 年 1 月から平成 30 年 1 月までの当院総合診療科での約 2 万件のカルテ記載データを用いて、辞書・言語モデルの拡充を行った。

2) 機材

- i). ピンマイク (audio-technica AT9901) + IC レコーダー (SONY ICD-SX2000)
- ii). マイクフォンアレー (ReSpeaker Mic Array v2.0) + Audacity (Windows 版)

3) 辞書・言語モデルの構築

3.1) 書き起こしテキスト

書き起こしテキストを形態素解析 (解析エンジン: MeCab + 解析辞書: UniDic) した。解析辞書には上述の薬剤名辞書の項目、および、頻出する未知語を追加した。

3.2) カルテ記載テキスト

カルテ記載テキストを形態素解析 (解析エンジン: MeCab + 解析辞書: UniDic) した。解析辞書には上述の薬剤名辞書の項目を追加した。個人情報に当たると思われる人名・地名は解析結果中ですべて伏せ字に変換した。

3.3) 辞書・言語モデル

音声認識ツールキット Kaldi を用いて、『日本語話し言葉コーパス』(Corpus of Spontaneous Japanese : CSJ)²⁾および上

記書き起こしテキスト・カルテ記載テキストのいくつかの組み合わせについて、辞書・言語モデルを構築した。

4) 辞書・言語モデル/音響モデルの評価

医療面接の音声認識における認識精度において、本研究で構築した辞書・言語モデルならびに音響モデルを、既存の CSJ に基づく辞書・言語モデルを用いた場合の音声認識率と比較して評価した。

音響モデルの構築には Kaldi5.5³⁾を用い、リアルタイム音声認識には pykaldi⁴⁾を用いた。Kaldi は最新のアルゴリズムを積極的に採用することで他のオープンソース音声認識エンジンと比較し、一般に非常に高い認識率を示す⁵⁾。

音声は使用した IC レコーダーの仕様上、44.1 kHz/16 bit/2 ch で集音されているため、音声認識に当たって sox コマンドにて 16 kHz/16 bit/ 1 ch へ変換を行った。

認識率の算出方法にはいくつかの定義が存在するが、本検証では雑音等による挿入誤りが一定程度含まれることが想定されるため、これを排除した単語誤り率 Word Error Rate (WER:%) を評価指標とした(式 1)。また、WER に加えて文字誤り率 Character Error Rate (CER:%) も評価指標として採用した(式 2)。CER を用いる理由は、日本語の解析において単語分割の曖昧性が存在するためである。

WER (%) =

$$\frac{\text{挿入単語数} + \text{置換単語数} + \text{削除単語数}}{\text{正解単語数}} \times 100 \dots$$

(式 1)

Correct (%) =

$$\frac{\text{挿入文字数} + \text{置換文字数} + \text{削除文字数}}{\text{正解単語数}} \times 100 \dots$$

(式 2)

4.1) 音響モデルに関する比較

下記の音響モデルによる音声認識率の違いを比較した。

- ① CSJ のみ
- ② CSJ + 収録音声データ
- ③ 収録音声データのみ

このとき、言語モデルには「CSJ(約 240 時間分) + 書き起こしテキスト(ピンマイクによる医師の発話データ、約 40 時間分)」を、音声認識率の検証用データには、各種モデル構築に使用していない、「収録音声データ(ピンマイクによる医師の発話データ、74:18 分、10,606 単語)」を、それぞれ用いた。

4.2) 言語モデルに関する比較

下記の言語モデルによる音声認識率の違いを比較した。

- ④ CSJ のみ
- ⑤ CSJ + 書き起こしテキスト
- ⑥ CSJ + 書き起こしテキスト + カルテ記載テキスト
- ⑦ 書き起こしテキストのみ

このとき、4.1)の結果をふまえて、音響モデルには「収録音声データ(ピンマイク

による医師の発話データ)のみ」を、音声認識率の検証用データには 4.1)と同様の医師の発話データを、それぞれ用いた。

4.3) 医師および患者の発話に関する比較

音声認識率の検証用データとして、各種モデル構築に使用していない、ピンマイクによる患者の発話データ(108:23 分、10,844 単語)を用いた。このとき、言語モデルには「CSJ(約 240 時間分) + 書き起こしテキスト(ピンマイクによる患者の発話データ、18 時間分)」を、音響モデルには「CSJ(約 240 時間分) + 収録音声データ(ピンマイクによる患者の発話データ、18 時間分)」を、それぞれ用いた。

4.4) 収録機材に関する比較

収録機材としてピンマイクを用いた場合とマイクロフォンアレーを用いた場合とで音声認識率を比較した。マイクロフォンアレーの音声認識率を検証する際には、言語モデルを「CSJ(約 240 時間分) + 書き起こしテキスト(マイクロフォンアレーによる発話データ、13 時間分)」、音響モデルを「CSJ(約 240 時間分) + 収録音声データ(マイクロフォンアレーによる発話声データ、13 時間分)」を用いて構築し、音声認識率の検証用データには、マイクロフォンアレーによる収録音声データ(34:12 分、7,014 単語)を用いた。

4.5) 書き起こしテキストの読み上げによる

音声認識率の検証

医療面接時の書き起こしテキストを元に、実際の運用環境に近い条件(生活雑音程度の環境ノイズ、胸部付近に取り付けたピンマイク、医療面接程度の声量)において、明瞭に発話することを意識して音声を再収録し、音声認識率を検証した。このとき、言語モデルには「CSJ(約240 時間分) + 書き起こしテキスト(ピンマイクによる医師の発話データ、約40 時間分)」を、音響モデルには「CSJ + 収録音声データ(ピンマイクによる医師の発話データ、約40 時間分)」を、それぞれ用いた。具体的には、4.2)で用いた検証用データの一部を取り出し、その音声認識率を算出した結果と、その発話データに該当する書き起こしテキストを改めて読み上げ、音声認識率を算出した結果とを比較した。

<身体所見記載に関する辞書編纂>

1) 辞書編纂

一般財団法人医療情報システム開発センター(以下、MEDIS-DC)のホームページにおいて公開されている「症状所見マスター<身体所見編>(2014.03.06更新)」⁶⁾を用いて、実地医家で頻用されと思われる用語を選定し、音声認識を目的とした辞書編纂を行った。このときに行った手順を具体的に記録し、今後の効率的な辞書編纂の参考となるようにした。

具体的な辞書編纂手法は以下のとおり。

- ① 症状所見マスターに含まれる身体所見用語約 3,550 項目から実地医家において頻用されると想定される用語を手作業で選定し、約740 項目とした。
- ② 症状所見マスターの「基本用語」(身体所見用語)を形態素解析(解析エンジン:MeCab + 解析辞書:UniDic)し、発音形を得た。誤りは手作業で修正した。
- ③ 症状所見マスターの「値」を精査し、必要に応じて「値」の選択肢を拡充(「あり/なし」に「状態/傾向」を追加するなど)し、「判定用語」を作成した。
- ④ ②の身体所見用語と③の判定用語を組み合わせ、「傾眠 + あり/なし/状態/傾向」「体温 + Number」のような身体所見パターンの一覧を作成した。なお、「視力低下」のように冒頭に「左・右・左側・右側・両側」が付くパターンや、「体温 + Number」のように末尾に単位(「°C」など)が付くパターンにも対応した。
- ⑤ ④の身体所見パターンから音声認識システム Julius の記述文法(grammar と voca)に変換した。

2) 辞書の評価

(1) 身体所見用語の認識

編纂した身体所見用語辞書の有効性を評価するために、身体所見用語のみの音声認識における身体所見用語辞書の効果を検証した。検証方法は、辞書編纂

時に選定した用語から 200 件をランダムに抽出し、3名の医師がそれぞれ読み上げ、その認識精度を測定した。音声収録は薬剤名辞書の評価の場合と同様に、2台の Windows 端末に、i) 口元にセットした単一指向性マイク(サンワサプライ MM-MCU02BK)および ii) マイクロフォンアレー(ReSpeaker Mic Array v2.0)をそれぞれ接続して同時に行い、音声認識エンジンには Julius を用いた。なお、評価については Julius に付属している標準辞書のみを用いた場合と標準辞書に身体所見用辞書を加えた場合とで比較し、音素列での一致度(ゲシュタルトパターンマッチング)によって認識精度の比較を行った。

(2) 身体所見パターンの認識

身体所見用語の後に「あり／なし」などの判定用語を続けた身体所見パターンに対する認識精度を検証した。身体所見パターンを辞書に登録したものからランダムに 200 件作成し、それを 3名の医師が読み上げ、精度を評価した。音声収録も(1)と同様に口元マイクおよびマイクロフォンアレーを用いて行った。

(3) より一般的な身体所見記述の認識

実際の診療録の作成時には、「両側下腿に圧痕性浮腫を認める」のようなより一般的な記述を用いることがある((2)の身体所見パターンでは「圧痕性浮腫あり」)。このような、より一般的な身体所見記述の認識を、<医療面接の音声自動テキ

スト化の精度向上に関する分析>において使用した連続音声認識システムを用いて試み、その精度を検証した。

音声収録は(1)と同様に i)口元マイクおよび ii)マイクロフォンアレーを用いて行い、次の記述例を読み上げた。「体温: 36.8℃、サチュレーション:98%、脈拍: 88/分、血圧:140/79。眼瞼結膜貧血なし、眼球結膜黄疸なし・充血なし。側頭動脈拍動 触知良好、索状物触知なし。舌表面びらん形成あり 粘膜疹なし 咽頭後壁発赤なし 頸部リンパ節腫脹なし。呼吸音清、心音整・心雑音なし 腹部平坦・軟、圧痛なし 触診で肝脾腫なし。上肢・下肢 ファーレン徴候陰性、チネル徴候陰性。触覚および痛覚に異常なし 振動覚(右 11 秒、左 11 秒) 位置覚正常。深部腱反射:膝蓋腱(左右とも亢進減弱なし)、アキレス腱(左右とも亢進減弱なし)、上腕二頭筋(左右とも亢進減弱なし)、上腕三頭筋(左右とも亢進減弱なし)、腕橈骨筋(左右とも亢進減弱なし)、オトガイ反射(左右とも亢進減弱なし)。ラゼーグ徴候陰性 バビンスキ徴候陰性。膝関節腫脹なし、圧痛なし、自動時痛なし、他動時痛なし。股関節腫脹なし、圧痛なし、自動時痛なし、回旋時痛なし。」

使用した言語モデルおよび音響モデルは次のとおりである。

- i) 口元マイクを用いた場合

言語モデル:「CSJ(約 240 時間分) + 書き起こしテキスト(ピンマイクによる医師の発話データ、約 40 時間分)」

音響モデル:「収録音声データ(ピンマイクによる医師の発話データ)のみ」

ii) マイクロフォンアレーを用いた場合

言語モデル:「CSJ(約 240 時間分) + 書き起こしテキスト(マイクロフォンアレーによる発話データ、13 時間分)」

音響モデル:「CSJ(約 240 時間分) + 収録音声データ(マイクロフォンアレーによる発話声データ、13 時間分)」

(倫理面への配慮)

本研究は、千葉大学大学院医学研究院倫理審査委員会、および千葉大学大学院医学研究院利益相反委員会の承認を得て行った。具体的には、臨床研究に関する指針、個人情報保護法、医療・介護関係事業者における個人情報の適切な取り扱いのためのガイドラインおよび医療情報システムの安全管理に関するガイドライン(第 4.3 版)に基づき施行した。

音声収録に際しては患者からの同意を必須としており、本研究への協力の有無は患者の自由意思であること、また研究に協力しないことで患者が一切の不利益を被らないことを担当医師から口頭および文書で説明した。なお、音声収録データは辞書編纂の目的にのみ使用し、

診療記録等のその他の目的で使用することは一切ないことを明示した。

過去の診療録データの利用にあたっては、音声収録の対象者とは別個で、過去に受診した患者データを利用するため、改めて本人から同意を得ることが困難であるが、情報のみを利用する研究であるので、病院の掲示板等に本研究に関する掲示文を掲載し、オプトアウト手続きによって利用することとした。

C. 研究結果

<音声コマンドに関する辞書編纂>

1) 電子カルテ項目名の標準呼称案の作成

主要電子カルテベンダーのうち回答を得られた4社の電子カルテに存在する機能・項目名を整理したところ、各ベンダーから標準的機能として提供された機能・項目数は約 120~200 項目であり、それらのうち同一機能として多くのベンダー間で共通して存在している機能は約 50 項目抽出された。それらを大きくグループ分けすると、オーダー系、検査結果系、参照系、医師記録系、看護系、薬剤系、病棟系、病床管理系などにまとめられた。これらの各項目に対して、一般的と考えられる呼称を標準呼称案として付与した(表1)。

2) 薬剤名辞書

(1) 辞書編纂

「B. 研究方法」で述べた手法で 2553 語からなる薬剤名辞書を編纂した。辞書エントリーの例を表2に示す。音声認識用の音素列は「発音形」から生成した。

(2) 辞書の評価

音声認識精度における薬剤名辞書の効果を検証した結果を表3、図1に示す。

i) 口元マイクを用いた場合

医師 A および医師 B では、標準辞書を用いた場合に比して薬剤名辞書を用いた場合の正答率がそれぞれ 48.62%、46.34%向上していた。医師 C では薬剤名辞書を用いた場合の正答率が改善していたが、その向上は 13.85%にとどまっていた。3 者の平均値で見ると薬剤名辞書を用いた場合、正答率の 36.27%の向上が得られた。

ii) マイクロフォンアレーを用いた場合

i)と同様に、医師 A、B、C のいずれでも、標準辞書を用いた場合に比して薬剤名辞書を用いた場合の正答率がそれぞれ 33.3%、40.25%、22.73%向上していた。3 者の平均値では薬剤名辞書を用いた場合、正答率の 32.1%の向上が得られた。

<医療面接の音声自動テキスト化の精度向上に関する分析>

4.1) 音響モデルに関する比較

各音響モデルでの WER、CER を表4に示す。CSJ を用いずに収録音声データのみを用いて構築した音響モデル

(③)の場合が WER、CER ともに小さい結果となった。

4.2) 言語モデルに関する比較

各言語モデルでの WER、CER を表5に示す。診察室内で収録された音声データはタイピング音や衣擦れ音等の環境音が比較的大きく、書き起こしテキストにも多少の表記揺れが存在する。そのため、全てのモデルを通じて WER、CER ともに比較的大きな値をとる結果となったが、各モデルによる認識結果を比較すると、言語モデルとして「CSJ+書き起こしテキスト+カルテ記載テキスト」を用いたモデル⑥が最も小さい WER、CER を示した。

なお、各言語モデルの構築に用いた元データに含まれる単語トークンの総数はそれぞれ下記のとおりであった。

CSJ:	7,494,708 語
書き起こし:	611,654 語
カルテ記載:	10,506,828 語

4.3) 医師および患者の発話に関する比較

結果を表6に示す。同様のピンマイクによる医師の発話の音声認識率と比較して、患者の発話では WER、CER ともに大きな値となった。

4.4) 収録機材に関する比較

マイクロフォンアレーの音声認識率を検証するため、言語モデルおよび音響モデルを「CSJ(約 240 時間分)+収録音声データ(マイクロフォンアレーによる

収録音声データ、13 時間分)」を用いて構築した。検証用データには、マイクロフォンアレーによる収録音声データ(34:12 分、7,014 単語)を用いた。結果を表7に示す。ピンマイクによる医師および患者の音声認識率と比して大きな WER、CER となった。

4.5) 書き起こしテキストの読み上げによる音声認識率の検証

4.2)で用いた検証用データのうち、6:13 分の収録音声データ(ピンマイクによる医師の発話データ)の音声認識率を算出した結果、WER = 65.58%、CER = 40.24%であった。この発話データに該当する書き起こしテキストを改めて読み上げ、その音声認識率を算出した結果、WER = 29.81%、CER = 18.54%となり、大幅な認識率の向上が見られた(表8)。

<身体所見記載に関する辞書編纂>

1) 辞書編纂

方法で示した手順により構築した辞書の具体的構成例を表9に示す。

2) 辞書の評価

(1) 身体所見用語の認識

身体所見用語のみの音声認識精度における身体所見用辞書の効果を検証した結果を表10、図2に示す。

i) 口元マイクを用いた場合

医師 A、B、C のいずれでも標準辞書を用いた場合には 40%前後の正答率だったが、身体所見用辞書を用いた場合には 66~78%への正答率の向上が見られ

た。3 者の平均値で見ると身体所見用辞書を用いた場合、標準辞書を用いた場合と比して正答率が 35.21%の向上が得られた。

ii) マイクロフォンアレーを用いた場合

i)と同様に、医師 A、B、C のいずれでも標準辞書を用いた場合には 40%前後の正答率であったが、身体所見用辞書を用いた場合には正答率は 69~76%に向上した。3 者の平均値では身体所見用辞書を用いた場合、正答率の 35.99%の向上が得られた。

この検証では口元マイクとマイクロフォンアレーとで認識精度に大きな差はなかった。

(2) 身体所見パターンの認識

身体所見パターンの音声認識精度の検証結果を表11、図3に示す。口元マイクではいずれの医師でも 70%前後の認識精度を示したが、一方、マイクロフォンアレーでは(1)の検証と比較して認識精度が安定せず、話者によっては約 30%の認識精度となった。

(3) より一般的な身体所見記述の認識

より一般的な身体所見記述の連続音声認識システムでの認識結果を表12に示す。口元マイクを用いた場合でも WER 67.39%、CER 46.69%であり、認識精度は低い結果となった。

D. 考察

<音声コマンドに関する辞書編纂>

1) 電子カルテ項目名の標準呼称案の作

成

各ベンダー間で共通している機能や項目を抽出したところ約 50 項目が抽出された。各ベンダーから標準的機能として提供された項目数は約 120～200 項目であり、共通していた約 50 項目以外の残りの多くの項目については、同じような機能を指していると思われる項目があるものの、各機能の差異を考慮すると同一のものとして標準呼称を付与することはできなかった。これらは電子カルテそのものの規格がまだ十分には標準化されておらず、各ベンダー間での互換性が乏しいこととも関連した問題であると考えられた。

今後、電子カルテ内の特定の機能や項目を音声で呼び出す技術の現場導入を推進するためには、それらの機能や項目に対して、電子カルテ利用者の多くがある程度共通した呼称を用いていることが求められる。したがって、電子カルテの標準化が進み、機能や項目の名称が標準化されることは音声認識技術の現場活用にとっても大きな意義があると考えられる。

2) 薬剤名辞書

音声認識率の検証では、被験者間において認識の正答率に大きな差があり、話す速度や話し方、声量などの要素によって正答率が変化するという、これまでに一般的に理解されているような結果となったが、いずれの医師、いずれの音声

入力条件下においても薬剤名辞書の追加による認識精度の向上が認められた。本研究で行った手順で編纂された薬剤名辞書により、実用的な音声認識結果が得られる可能性が示された。

また、一般には遠隔マイクよりも口元にマイクをセットした場合の認識精度が高くなる傾向にあるが、標準辞書を使用した場合には必ずしもそのような結果とはならなかった。薬剤名辞書を使用した場合には医師 A、B ではその傾向が見られたが、医師 C では逆に口元マイクの方で精度が低くなるという結果だった。この原因としてはマイクが近すぎて風切り音が発生した可能性や、音割れが起こった可能性が推察された。

<医療面接の音声自動テキスト化の精度向上に関する分析>

音響モデルに関しては、「CSJ+収録音声データ」の場合よりも、「収録音声データのみ」でモデル構築を行った場合に最も認識率が良い結果となった。一般にモデル構築に用いるデータ量が多いほど高い認識率を示すと予想されるが、今回の検証では、約 240 時間分の CSJ データよりも約 40 時間分の収録音声データから構築した音響モデルの方が高い認識率を示したことは注目すべき点である。CSJ の音声は比較的静穏な状況で接話マイクによる収録であるのに対し、今回診察室内で収録された音声は胸部付近に装着したピンマイクを使用してい

るため準遠隔音声認識であるとともに、タイピング音や衣擦れ音など比較的大きな室内雑音が重畳しており、音声の品質が必ずしも高くない。このような CSJ データと収録音声の性質の差が、認識率の違いとして現れたものと考えられる。

一方、言語モデルに関する比較の結果から、言語モデルについては認識率の向上の程度は小さいものの「CSJ+医療面接時の書き起こしテキスト+カルテ記載テキスト」を用いた場合において WER と CER が最も小さい結果となった。言語モデル構築に用いたいずれのデータも認識率の向上に寄与していると考えられたが、最も効果的だったのはカルテ記載テキストであった。この結果において、データに含まれる単語トークンの総数による影響は無視できない。言語モデル構築に用いた元データに含まれる語数を比較すると、CSJ では約 750 万語、書き起こしテキストでは約 61 万語、カルテ記載テキストでは約 1,050 万語となっており、元データに含まれる語数の差が大きかった。つまり、カルテ記載テキストは CSJ よりも語数が多かったため、その追加による認識率の改善効果が大きくなった可能性がある。しかしながら、この結果は、「CSJ+医療面接時の書き起こしテキスト」を汎用モデルとし、そこに特定領域のテキストデータ(今回は総合診療科カルテ記載テキスト)を追加することによって、連続音声認識をその領域によ

り適応させることができた、ということを示している。また、モデル⑦は④の場合と比べて同等以上の認識精度を示しており、CSJ と書き起こしテキストのそれぞれに含まれる語数の差を考慮すると、注目に値する。書き起こしテキストは他の元データに比して著しく語数が少なかったため、それを補正すればより明確な改善効果が得られた可能性がある。例えば、対処法として、書き起こしテキストに実際に登場した語を単純に N 倍(例えば N=2, 5, 10 など)に「水増し」して擬似的に語数を増やした後、言語モデル構築に利用するなどの処理を行えば、より良い結果が得られる可能性がある。N をいくつに設定するのが適切かは帰納的に定める必要があり、今後の課題である。

医師の発話と患者の発話の認識率を比較した場合、後者の場合の認識率が大幅に低かった。これには、音響モデルの構築に用いたピンマイクによる患者発話のデータ量(約 18 時間分)が、医師の場合のそれ(約 40 時間分)よりも少なかったことと、一般に医師に比べ患者の発話では言葉が不明瞭になるケースが多いことが影響しているものと推察された。

収録機材による比較では、ピンマイクの場合に比して、マイクロフォンアレーの場合で認識精度が大きく劣る結果となった。マイクロフォンアレーは内部に複数のマイクを搭載しており、ビームフォーミングにより音の発生方向の感度を高める

ように指向性を変化させることのできるマイクである。このマイクロフォンアレーは、一般に遠隔音声認識や騒音下音声認識に適しており、遠隔音声認識の精度向上に寄与することが知られているが、今回の評価対象ではビームフォーミング技術よりも発話源にマイクをできるだけ近づけることの方がより有利であったためと考えられる。また、ピンマイクの場合に比して、マイクロフォンアレーの収録音声データ量は約 13 時間分と少なかったこともその原因として考えられた。

書き起こしテキストを明瞭に発話することを意識して読み上げることで、大幅な認識率の向上が得られたことから、発話の際の意識の違いも認識精度に大きな影響を与えることが示された。明瞭な発話については患者側にこれを負担させることは困難であることが予想されるが、医師側においては意識的に明瞭な発話を心がけることで実用的な音声認識精度が得られる可能性がある。

医療面接時の音声自動テキスト化の精度向上に関する今後の対策として、次の 4 点が挙げられる。1. より一層の言語モデル・音響モデルの改良、2. 歪みの少ないノイズ低減処理、3. 室内残響の低減、4. Voice Activity Detection(VAD)の高度化、である。具体的には、1.では今後より多くの実データ(診察室内音声および書き起こし)やカルテ記載などの領域関連テキストデータを蓄積していくこ

と、2.では Denoising Auto Encoder 等を用いたノイズ低減、3.では診察室内のインパルス応答を測定し、それを畳み込むことでクリーン音声(CSJ 等)を実データに近づけ音響モデルを構築すること、4.では発話とタイピング音等のノイズのより高度な弁別を図ることでより高い精度での音声認識が可能になるものと考えられた。

<身体所見記載に関する辞書編纂>

身体所見用語のみの音声認識率の検証では、いずれの医師、いずれの音声入力条件下においても身体所見用辞書の追加による認識精度の向上が認められた。今回編纂した辞書では身体所見用語は約 740 項目に限られているが、用語と判定用語を組み合わせて登録する方法により、収載された用語であれば実用的なレベルで認識されることが実証できた。また、身体所見パターンの認識でもほぼ同程度の精度が得られた。

このことから、今後は身体所見記載に用いられる用語を拡充すること、それらに付随する判定用語を適切に用意することが有益であると考えられる。しかし、身体所見記載で用いられる用語も標準化されているとは言えず、加えて日本語表現ではなく英語などの外国語表現を好んで用いる医療従事者もいることから、これらを全て網羅するには、実際にカルテに記載された身体所見データを収集するなどの別の対策が必要となるも

のと推察される。また、一般に孤立単語認識は比較的lowコストで実用的な精度が得られる傾向にあるが、実運用の場面では単語のみでの発音には一定の慣れが必要で、不必要な間投詞や文末表現を付けてしまったり、途中で発話が途切れてしまったりして、認識精度が低下する危険性がある。そのため、実際の利用場面を想定すると連続音声認識の方が望ましい可能性がある。今回の簡単な検証では連続音声認識システムでの認識率が低かったが、これは身体所見の記載そのものが、医療面接の発話内容と比べて前後の脈絡のない、個々に独立した記載内容であるためと推察された。特定領域への適応という観点では、実際に身体所見として記載されたテキストデータなどを収集して言語モデルの構築に活用するなどの対策が有効と推察される。

より一般的な身体所見記述にも対応するための今後の対策をまとめると、①孤立単語認識を採用する場合、バリエーションを数多く網羅して、パターン登録する、②連続音声認識を採用する場合、この場面に合った統計モデルを構築する、③孤立単語認識と連続音声認識を並行して実施し、各部分単語列に対して、より蓋然性の高い方を採用する処理を実施する、などが考えられる。

本研究で構築した辞書、言語モデル、音響モデルについては、下記ホームページ上での配布を予定している。

千葉大学医学部附属病院企画情報部
ホームページ

<https://www.ho.chiba-u.ac.jp/section/ikaku/index.html>

引用文献

- 1) 野田和敬, 生坂政臣, 傳康晴, 鈴木隆弘, 大平善之, 上原孝紀, 島井健一郎. 電子カルテと連携する音声認識システムに関するニーズ調査. 医療情報学 2018: 38(Suppl.); 316-320.
- 2) https://pj.ninjal.ac.jp/corpus_center/csj/ (2019/3/30 アクセス)
- 3) <http://kaldi-asr.org/> (2020/3/30 アクセス)
- 4) <https://github.com/pykaldi/pykaldi> (2020/3/30 アクセス)
- 5) C. Gaida, P. Lange, R. Petrick, P. Proba, A. Malatawy, and Suendermann-Oeft, Comparing open-source speech recognition toolkits, Tech. Rep., DHBW Stuttgart, 2014.
- 6) <https://www2.medis.or.jp/master/syoken/> (2020/3/30 アクセス)

E. 結論

電子カルテ項目名辞書、薬剤名辞書、身体所見用辞書はマスター等からの具体的な処理手順を整理し、それらによる

有効性を実証した。医療面接の音声自動テキスト化については実発話データおよび特定領域に関連したテキストデータを集積させることが重要である。

F. 健康危険情報

特記事項なし。

G. 研究発表

1) 論文発表

特記事項なし。

2) 学会発表

1) 野田和敬, 生坂政臣, 傳康晴, 鈴木隆弘, 大平善之, 上原孝紀, 島井健一郎. 千葉県内臨床研修病院および診療所における電子カルテと連携する音声認識システムに関するニーズ調査. 第39回医療情報学連合大会. 2019/11/24, 千葉。

H. 知的財産権の出願・登録状況

特記事項なし。

表1) 電子カルテ機能・項目名の標準呼称案

系統	標準呼称案(ベンダー間)	よみがな
医師系	カルテ記載	カルテキサイ
	文書一覧	ブンショイチラン
	診療情報提供書作成／文書作成	シンリョウジョウホウテイキョウショサクセイ／ブンショサクセイ
	院内紹介	インナイショウカイ
	スキャン取込	スキャントリコミ
オーダー系	検体検査オーダー	ケンタイケンサオーダー
	細菌検査オーダー	サイキンケンサオーダー
	生理検査オーダー	セイリケンサオーダー
	病理検査オーダー	ビョウリケンサオーダー
	放射線検査オーダー	ホウシャセンケンサオーダー
	内視鏡オーダー	ナイシキョウオーダー
	処方オーダー	ショホウオーダー
	注射オーダー	チュウシャオーダー
	手術オーダー	シュジュツオーダー
	処置オーダー	ショチオーダー
	透析オーダー	トウセキオーダー
	食事オーダー	ショクジオーダー
	栄養指導オーダー	エイヨウシドウオーダー
	服薬指導オーダー	フクヤクシドウオーダー
	輸血オーダー	ユケツオーダー
予約オーダー	ヨyakオーダー	
検査結果系	画像参照	ガゾウサンショウ
	検体検査結果	ケンタイケンサケッカ
	細菌検査結果	サイキンケンサケッカ
	心電図参照	シンデンズサンショウ
	病理検査結果	ビョウリケンサケッカ
看護系	看護データベース	カンゴデータベース
	看護オーダー	カンゴオーダー
	看護サマリー	カンゴサマリー

	看護記録	カンゴキログ
	看護計画	カンゴケイカク
	看護必要度	カンゴヒツヨウド
	受持看護師登録	ウケモチカンゴシトウロク
病棟系	主治医登録	シュジイトウロク
	病名登録	ビョウメイトウロク
	DPC 入力	ディーピーシーニューリョク
	指導料・管理料	シドウリョウカンリリョウ
病床管理系	退院決定	タイインケツテイ
	退院サマリー	タイインサマリー
	退院サマリー一覧	タイインサマリーイチラン
	転室・転床	テンシツテンショウ
	入院決定	ニューインケツテイ
	入院予約	ニューインヨヤク
	外出・外泊登録	ガイシュツガイハクトウロク
薬剤系	持参薬入力	ジサンヤクニューリョク
	薬歴参照	ヤクレキサンショウ
	医薬品情報参照／DI 参照	イヤクヒンジョウホウサンショウ／ディーアイサンショウ
参照系	輸血歴参照	ユケツレキサンショウ
	手術歴	シュジュツレキ
	予約照会	ヨヤクショウカイ

表2) 医薬品データベースから変換した薬剤名認識用辞書の例

漢字名(正式)	カナ	書字形	発音形	オーダコード
IFN β 注 (300 万 IU/V)	IFN ベータ	IFN β	アイエフエヌベータ	N70101
クラシエ甘草湯エキス 細粒 3g/包	カンゾウトウ	甘草湯	カンゾートー	N01500
ケフラーカプセル 250mg	ケフラーカプセル 250mg	ケフラー	ケフラー	N00256,N00257
ケフラー細粒小児 用 100mg カ価/g	ケフラーサイリュウショウ ニヨウ100mg			

表3) 標準辞書および薬剤名辞書追加での音声入力機器別認識精度の比較(正答率(%))

i) 口元マイク				
	医師 A	医師 B	医師 C	平均値
標準辞書	35.73	44.43	33.68	37.95
薬剤名辞書	84.34	90.76	47.53	74.21
ii) マイクフォンアレー				
	医師 A	医師 B	医師 C	平均値
標準辞書	39.77	42.26	35.339	39.12
薬剤名辞書	73.09	82.51	58.05	71.22

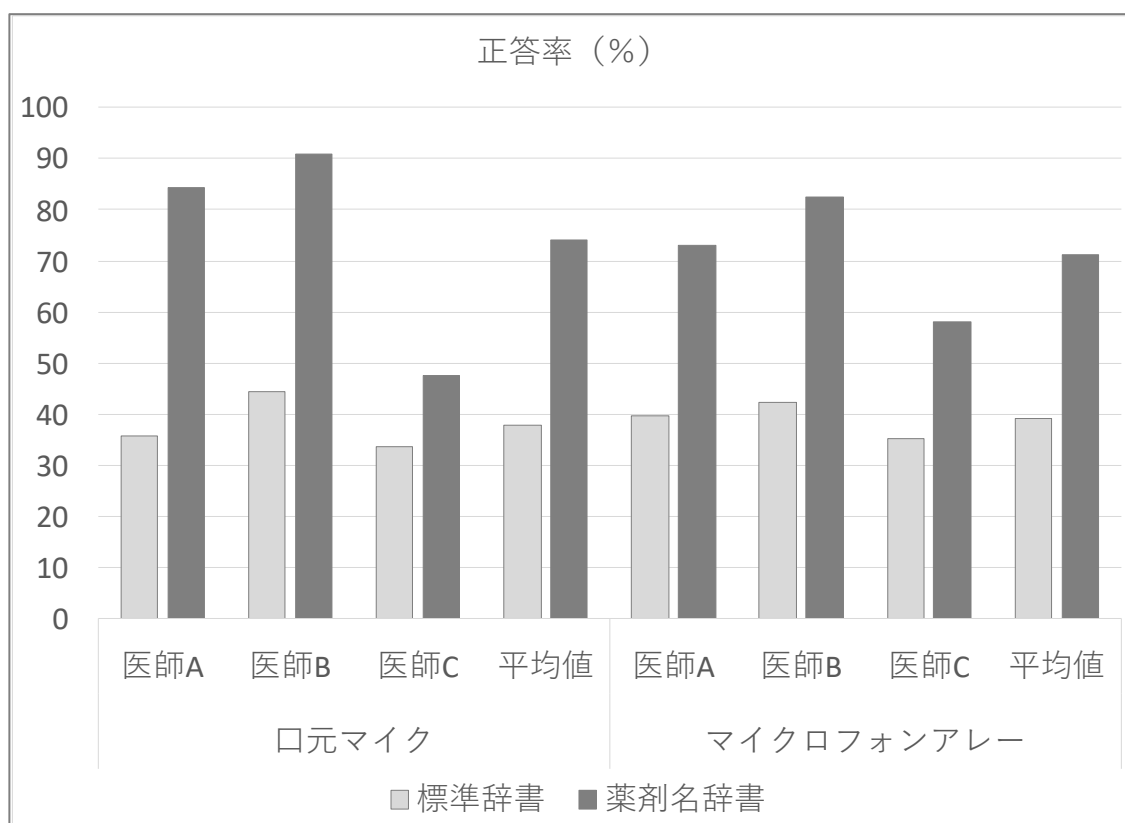


図1) 標準辞書および薬剤名辞書追加での音声入力機器別認識精度の比較(正答率(%))

表4) 音響モデルに関する音声認識率(WER、CER)の比較

Model			認識精度指標(%)		①との差	
No.	LM および辞書	AM	WER	CER	dWER	dCER
①*	CSJ+書き起こしテキスト* ¹	CSJ	72.33	50.35	—	—
②*	CSJ+書き起こしテキスト* ¹	CSJ+収録音声データ* ¹	72.51	50.44	0.18	0.09
③*	CSJ+書き起こしテキスト* ¹	収録音声データ* ¹	72.26	50.29	-0.07	-0.06

WER: 単語誤り率、CER: 文字誤り率、LM: 言語モデル、AM: 音響モデル、CSJ: 日本語話し言葉コーパス

※ 音声認識率の検証用データには、各モデルの構築に利用されていない、ピンマイクによる医師の発話データ(74:18分、10,606単語)を用いた。

*1 診察室での医療面接時にピンマイクを用いて医師の発話を収録(約40時間分)

表5) 言語モデルに関する音声認識率(WER、CER)の比較

Model			認識精度指標(%)		表4の①との差	
No.	LM および辞書	AM	WER	CER	dWER	dCER
④*	CSJ	収録音声データ* ¹	72.44	50.40	0.11	0.05
⑤* (=③)	CSJ+書き起こしテキスト* ¹	収録音声データ* ¹	72.26	50.29	-0.07	-0.06
⑥*	CSJ+書き起こしテキスト* ¹ +カルテ記載テキスト* ²	収録音声データ* ¹	67.94	47.62	-4.39	-2.73
⑦*	書き起こしテキスト* ¹	収録音声データ* ¹	72.43	50.40	0.10	0.05

WER: 単語誤り率、CER: 文字誤り率、LM: 言語モデル、AM: 音響モデル、CSJ: 日本語話し言葉コーパス

※ 音声認識率の検証用データには、各モデルの構築に利用されていない、ピンマイクによる医師の発話データ(74:18分、10,606単語)を用いた。

*1 診察室での医療面接時にピンマイクを用いて医師の発話を収録(約40時間分)

*2 総合診療科でのカルテ記載テキスト(約2万件)

表6) 医師および患者の発話に関する音声認識率(WER、CER)の比較

Model			認識精度指標(%)		表4の②との差	
No.	LM および辞書	AM	WER	CER	dWER	dCER
② ^{*1}	CSJ+書き起こしテキスト ^{*1}	CSJ+収録音声データ ^{*1}	72.51	50.44	—	—
⑧ ^{*2}	CSJ+書き起こしテキスト ^{*2}	CSJ+収録音声データ ^{*2}	81.67	57.93	9.16	7.49

WER: 単語誤り率、CER: 文字誤り率、LM: 言語モデル、AM: 音響モデル、CSJ: 日本語話し言葉コーパス

※1 音声認識率の検証用データには、各モデルの構築に利用されていない、ピンマイクによる医師の発話データ(74:18分、10,606単語)を用いた。

※2 音声認識率の検証用データには、各モデルの構築に利用されていない、ピンマイクによる患者の発話データ(108:23分、10,844単語)を用いた。

*1 診察室での医療面接時にピンマイクを用いて医師の発話を収録(約40時間分)

*2 診察室での医療面接時にピンマイクを用いて患者の発話を収録(約18時間分)

表7) 収録機材に関する音声認識率(WER、CER)の比較

Model			認識精度指標(%)		表4の②との差	
No.	LM および辞書	AM	WER	CER	dWER	dCER
② ^{*1}	CSJ+書き起こしテキスト ^{*1}	CSJ+収録音声データ ^{*1}	72.51	50.44	—	—
⑨ ^{*2}	CSJ+書き起こしテキスト ^{*2}	CSJ+収録音声データ ^{*2}	92.02	78.17	19.51	27.73

WER: 単語誤り率、CER: 文字誤り率、LM: 言語モデル、AM: 音響モデル、CSJ: 日本語話し言葉コーパス

※1 音声認識率の検証用データには、各モデルの構築に利用されていない、ピンマイクによる医師の発話データ(74:18分、10,606単語)を用いた。

※2 音声認識率の検証用データには、各モデルの構築に利用されていない、マイクロフォンアレーによる医師および患者の発話データ(34:12分、7,014単語)を用いた。

*1 診察室での医療面接時にピンマイクを用いて医師の発話を収録(約40時間分)

*2 診察室での医療面接時にマイクロフォンアレーを用いて医師および患者の発話を収録(約13時間分)

表8) 書き起こしテキストの読み上げによる音声認識率の検証

Model			認識精度指標 (%)		⑩との差	
No.	LM および辞書	AM	WER	CER	dWER	dCER
⑩ ^{※1}	CSJ+書き起こしテキスト ^{*1}	CSJ+収録音声データ ^{*1}	65.58	40.24	—	—
⑪ ^{※2}	CSJ+書き起こしテキスト ^{*1}	CSJ+収録音声データ ^{*1}	29.81	18.54	-35.77	-21.70

WER: 単語誤り率、CER: 文字誤り率、LM: 言語モデル、AM: 音響モデル、CSJ: 日本語話し言葉コーパス

※1 音声認識率の検証用データには、各モデルの構築に利用されていない、ピンマイクによる医師の発話データ(06:13分)を用いた。

※2 音声認識率の検証用データには、※1の書き起こしテキストを改めて読み上げた発話データ(ピンマイクにより収録)を用いた。

*1 診察室での医療面接時にピンマイクを用いて医師の発話を収録(約40時間分)

表9) 身体所見用辞書の構成例

n.orth	n.pron	v.orth	v.pron	group
意識清明	イシキセーメー			
傾眠	ケーミン	あり,なし,状態,傾向	アリ,ナシ,ジョータイ,ケーコー	Ari_Nashi_Jotai
血圧	ケツアツ	Number_No_Number		Number_No_Number
体温	タイオン	Number		Number
心音	シンオン	正常,異常,陰性,陽性,良好, 異常なし	セージョー,イジョー,インセー,ヨーセー, リョーコー,イジョーナシ	Seijo_Insei
湿性ラ音	シッセーラオン	あり,なし,聴取,聴取しない	アリ,ナシ,チョーシュ,チョーシュシナイ	Ari_Nashi_Choshu

発話認識パターン例

- group 列が空 → n.pron のみで発話
- v.pron 列が空 → group でパターン分け
 - Number_No_Number: 「Number の Number」
 - Number: 一般的な数値 (小数点ありなしに対応)

表10) 標準辞書および身体所見用辞書追加での身体所見用語に対する音声入力機器別認識精度の比較(正答率(%))

i) 口元マイク				
	医師 A	医師 B	医師 C	平均値
標準辞書	39.52	40.03	37.27	38.94
身体所見用辞書	78.28	77.83	66.35	74.15
ii) マイクフォンアレー				
	医師 A	医師 B	医師 C	平均値
標準辞書	36.41	40.48	37.36	38.08
身体所見用辞書	75.86	76.94	69.40	74.07

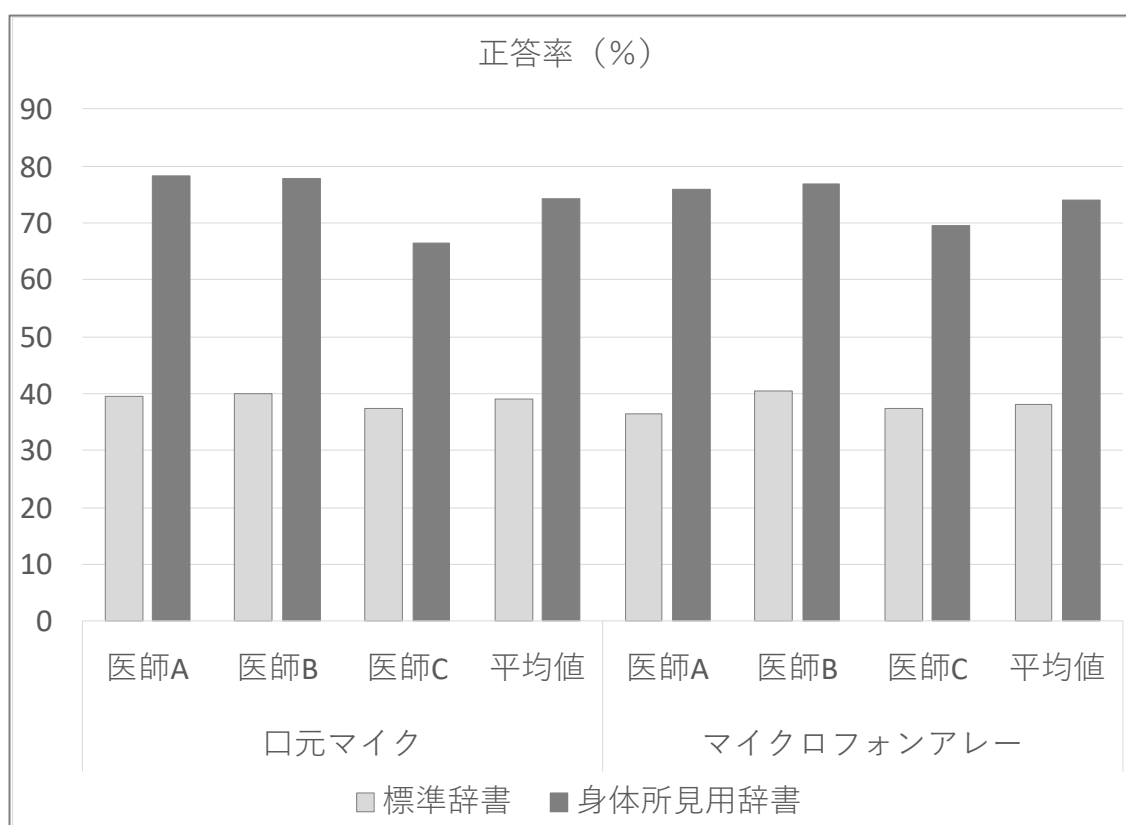


図2) 標準辞書および身体所見用辞書追加での身体所見用語に対する音声入力機器別認識精度の比較(正答率(%))

表11) 身体所見用辞書での身体所見パターンに対する音声入力機器別認識精度の比較(正答率(%))

iii) 口元マイク				
	医師 A	医師 B	医師 C	平均値
身体所見用辞書	71.73	69.85	72.62	71.40
iv) マイクフォンアレー				
	医師 A	医師 B	医師 C	平均値
身体所見用辞書	59.89	30.41	71.21	53.83

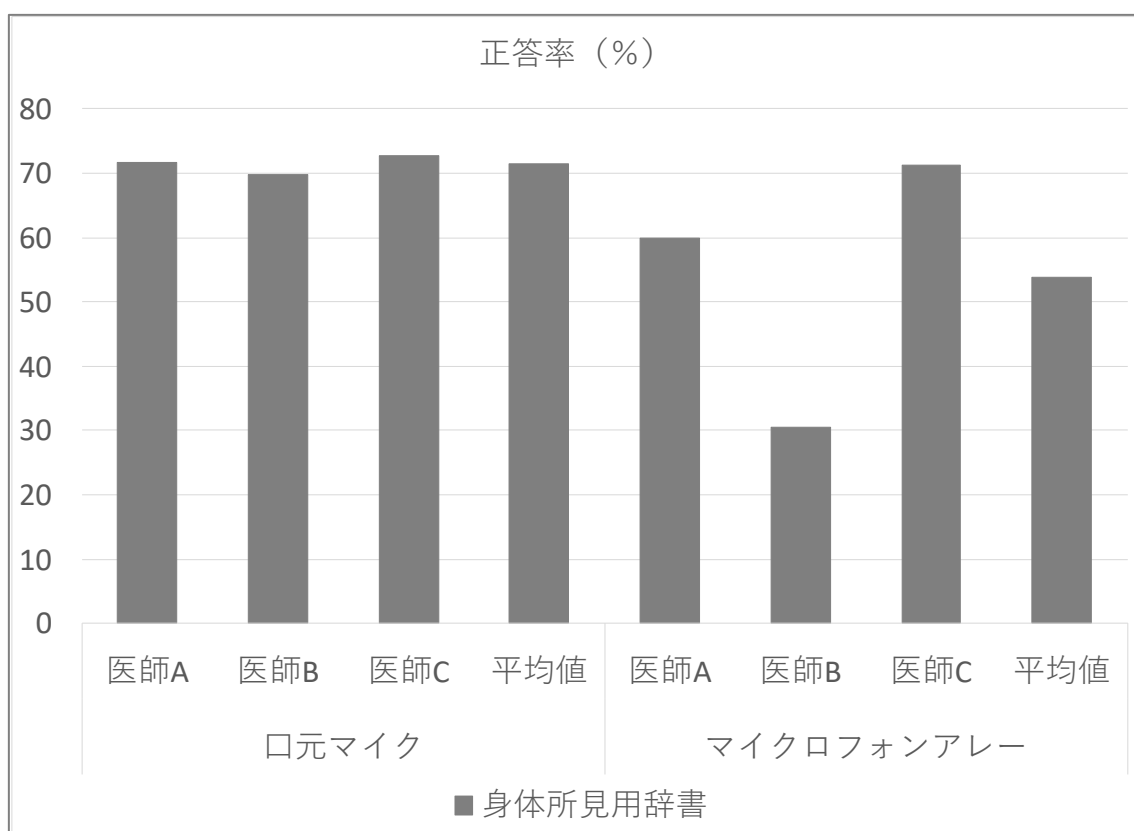


図3) 身体所見用辞書での身体所見パターンに対する音声入力機器別認識精度の比較(正答率(%))

表12) より一般的な身体所見記述の連続音声認識システムでの音声認識率の検証

検証条件				認識精度指標 (%)	
No.	収録機材	LM および辞書	AM	WER	CER
⑫	口元マイク	CSJ+書き起こしテキスト* ¹	収録音声データ* ¹	67.39	46.69
⑬	マイクロフォンアレー	CSJ+書き起こしテキスト* ²	CSJ+収録音声データ* ²	79.35	57.72

WER: 単語誤り率、CER: 文字誤り率、LM: 言語モデル、AM: 音響モデル、CSJ: 日本語話し言葉コーパス

*1 診察室での医療面接時にピンマイクを用いて医師の発話を収録(約 40 時間分)

*2 診察室での医療面接時にマイクロフォンアレーを用いて医師および患者の発話を収録(約 13 時間分)