

音声認識を活用した医療用電子カルテ項目別自動入力インタフェースの構築

前島 亮, 若林 佑幸, 北岡 教英
豊橋技術科学大学

Abstract: 本研究では、医療従事者が電子カルテの入力業務をより効率的に行うことを目的とし、音声による入力インタフェースを構築する。従来のキーボード入力に比べ、音声入力はより迅速かつ自然な入力方法であり、医療従事者の負担軽減につながると期待される。本研究では、実用的なシステムの実現に向け、電子カルテ作成のための音声入力インタフェースを提案する。提案するインタフェースでは、医療従事者の発話音声を確認した後、大規模言語モデルによって電子カルテの各入力項目に従い認識された文を整形することによって、効率的な電子カルテ入力を実現する。

1 はじめに

近年、音声認識の技術は幅広い分野で導入されるようになってきている。Siri や Alexa といったバーチャルアシスタントや AI スピーカーなどが代表的な例である。また最近ではコールセンターでのお客様対応や、テレビの字幕作成などでも導入されるようになり、ますます導入の幅を広げている。しかし日本の医療分野においては依然として導入されていないというのが現状である。すでにアメリカでは手術の現場で音声認識の技術が導入されており [1]、日本の医療分野における IT 化が遅れをとっていることがわかる。音声認識技術を医療分野に導入することで効率をあげることができるだけでなく、手動入力によるヒューマンエラーを未然に防ぐことも可能となる。また、現在の病院の課題として、回診時での電子カルテ入力の手間が挙げられる。回診中に電子カルテに入力を行うことは、患者の機嫌を損ねることにつながる。回診後に電子カルテ入力を行うためにはメモを取る必要があり、効率性、正確性ともに問題がある。

本稿では以降「音声によって入力される、項目別の入力を行うために必要な文」を「音声入力文」、「音声入力文をカルテ項目に分けること」を「構造化」とそれぞれ表現する。

本研究では、実用的なシステムの実現に向け、電子カルテ作成のための音声入力インタフェースを提案する。人間はキーボード入力よりも音声入力の方が速いという研究結果がある [2]。この事実を考えれば、選択肢からの選択や郵便番号からの住所自動入力といった効率化手段に加え、音声入力を組み合わせることで、入力の効率化をより一層進めることが可能だと考えられる。一方、多種多様なアプリケーションで項目別入力の機会は頻繁にあり、個人情報を要求するフォームや電子カルテはその代表的な例である。しかし、多数の項目を1つずつ音声入力するのは項目選択が必要となり効率が悪い。そこでいくつかの項目内容を1度に含む音声入力文から構造化し自動で項目別に分けることとする。このような構造化を伴う音声入力インタフェースを構築する場合、図1のような流れとなる。



図 1: 構造化を伴う音声入力インタフェースの流れ

より効率性、正確性の高い構造化を伴う音声入力インタフェースを構築するためには以下の要素を考慮する必要がある。

- 音声入力文の認識結果と構造化結果をどのように連携するか
- どのように構造化するか

本研究では上記 2 点を考慮した音声による項目別自動入力インタフェースの構築を図る。

2 システムの基本構成

本研究では、看護師が検温時もしくは検温後に音声によって電子カルテ項目を入力するシステムを実装する。このシステムは、音声入力文の認識結果を構造化し、構造化されたカルテ項目を手直しして、問題がなければ電子カルテシステムに送信することを想定している。システムは Web 上で動作し、デスクトップ端末やモバイル端末など、Web が動作する環境であればどこからでもアクセスすることが可能である。

本システムの利用の流れを図 2 に示す。看護師が音声

入力を行うと、音声入力文の認識結果と構造化結果が表示される、なお、音声入力文が入力され終わったタイミングで自動で構造化される。看護師は構造化結果を確認し、必要に応じて修正し電子カルテシステムに送信する。認識結果文を直接編集することも可能である。

本システムの音声認識部分には、Web Speech API [3] を用いる。Web Speech API は World Wide Web Consortium (W3C) が策定した API であり、ブラウザからストリーミング音声認識を利用することができる。

本システムの構造化サーバー部分には、ルールベースの構造化を行う疑似的なサーバーもしくは大規模言語モデルを用いる。ルールベースの構造化におけるプロセスは、予め設定された特定のルールに従って音声入力文の認識結果を解析し、それらを構造化するというものである。具体的には、「体調は」というフレーズがテキスト中に存在するとき、その文は体調に関連しているという規則を設ける。この規則に従って、文が体調に関連しているか否かを判断し、それに基づきデータを構造化する。大規模言語モデルとは大量のテキストデータとパラメータ数によって学習された機械学習モデルで質問応答や文章生成を可能にする。3.1 節と 3.2 節で提案する手法ではルールベースの構造化を行う疑似的なサーバーを用いる。ルールベースの構造化を行う疑似的なサーバーは、実際には通信処理は行わず、Web ブラウザ上で構造化する。

本システムのインタフェースを図 3 に示す。画面上部には音声入力文の認識結果が、画面下部には追加、編



図 3: システムのインタフェース

集、削除が可能な各カルテ項目の構造化結果がそれぞれ配置されている。カルテ項目には一般的な電子カルテで使用されるバイタルサインを採用する。なお、各カルテ項目は 3.1 節の提案手法を除き複数の結果を保持することができる。例えば図 3 のように、音声入力文に体温に関して複数回言及されていた場合、体温項目には複数の結果が表示される。

3 提案手法

本研究では音声入力文の構造化による項目別自動入力法として、医療従事者が電子カルテの入力を想定したインタフェース構築手法を 3 つ提案する。なお、3.1 節、3.2 節ではルールベースの構造化、3.3 節では大規模言語モデルによる構造化を採用する。

3.1 音声入力文終端挿入による変更同期手法

同じ項目の内容が複数回見つかった場合、構造化結果には音声入力文中の最も後方に位置する関連内容が入

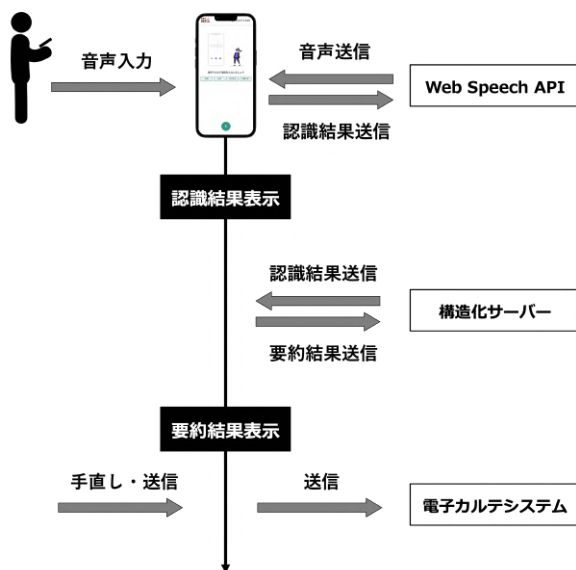


図 2: システム利用の流れ

る。例えば「体温 36.6 度で平熱です。あ、間違えた、体温は 36.8 度です」といった入力があった場合には体温の項目には「36.8 度」が入る。また、構造化後の項目を編集した際には、対応する文は音声入力文の最後に挿入される(図 4)。この仕様により、最初に入力される音声入力文は項目別の編集により変更されることはない。

3.2 参照による音声入力文・カルテ項目同期手法

構造化結果は音声入力文の一部とリンクして参照する。カルテ項目と音声入力文の状態が同期されるため、音声入力文、項目別のいずれかを編集・削除した際にもその変更が互いに直接反映される(図 4)。

3.3 大規模言語モデルを利用する手法

構造化サーバー部分に Chat Generative Pre-trained Transformer (ChatGPT) を用いる。ChatGPT とは、OpenAI が 2022 年 11 月に公開した人工知能チャットボットで大規模言語モデルである GPT-3 [4][5][6] を基に構築されている。今回は 2023 年 6 月現在利用可能である GPT-3.5 を使用する。

音声入力文の認識結果と電子カルテ項目をもとにプロンプトを作成し、ChatGPT に送信することによって、構造化を行う。図 5 にプロンプトと ChatGPT のレスポンス例を示す。プロンプトに JSON 形式のフォーマット命令を入れることによって、ChatGPT のレスポンスを JSON に変換しやすい形にしている。なお、受け取る JSON の各項目は全て配列を指定することにより、複数の構造化結果を保持することができるようになる。

この手法では 3.2 節の手法のように音声入力文とカルテ項目の同期は行わない。音声入力文の入力が終わった際、即時に ChatGPT へプロンプトの送信が行われ、正常に構造化結果が行われれば音声入力文内容はリセットされる。ユーザーの労力は最小限に抑えられるべきであるため [7]、構造化を行うためのボタンなどは配置せず、音声入力が終わったタイミングで自動で構造化までの処理が行われる。編集は各カルテ項目からのみ行うことを想定しており、その結果は音声入力文には反映されない。これは構造化が音声入力文の単純な参照ではなく、より高度な変更を伴うためである。図 5 のように、プロンプト経由でフォーマットのルールを設定することができる。具体的には文末を整えるというルールを設定することにより、本文の「体調は回復に向かっており」から体調項目に「回復に向かっており」を抽出するといったことが可能となる。

4 考察

3.1 節の提案手法は音声入力文に手を加えないため、音声入力文の情報を失わないという利点がある一方で、構造化結果を複数保持できず且つ構造化結果は音声入力文の後方が優先されるため、不本意な編集に気づかず電子カルテに送信してしまう可能性があるという欠点がある。加えて項目別の編集内容は音声入力文の末尾に挿入されるため、編集回数が増えると音声入力文が肥大化してしまうという問題がある。

3.2 節の提案手法では複数の構造化結果を所持することができるため構造化結果の確認を誤る可能性は低くなっている。項目別編集を行った際にも音声入力文に直接変更が反映されるため、音声入力文が肥大化するという問題もない。一方で音声入力文に変更が加えられるため、3.1 節の提案手法に比べて音声入力文の情報が失われてしまう可能性は高くなる。

3.3 節の提案手法では大規模言語モデルを用いることにより、高度な構造化を行うことができる。プロンプトを変更することによってカルテ項目や構造化のルールを容易に変更することができるため、ルールベースに比べて、開発コスト、柔軟性、拡張性に優れている。プロンプト経由で各カルテ項目のルールを反映することができるため、より安全なシステムを構築することができる。一方で大規模言語モデルを利用しているためにいくつかの制約が伴う。ルールベースの手法に比べて、大規模言語モデルによる構造化プロセスは時間を要するという欠点がある。特に構造化に高度な分析が必要な場合、更なる時間がかかる。さらに、この手法は音声入力文と構造化結果を同期させることが困難であるため、構造化過程において何らかの問題が生じた場合、その認識が難しいという課題がある。これは、提案システムの可用性と効率性を低下させる可能性がある。また、大規模言語モデルを構造化サーバーとして利用する際には、経済的な観点や個人情報の保護といった側面を慎重に考慮する必要がある。これは、言語モデルの利用料金や、ユーザーのプライバシーに関わる情報の取り扱いに関連している。したがって、これらの制約は本手法の採用における重要な課題となる。

5 結論

本研究では、医療従事者が電子カルテの入力業務をより効率的に行うための音声入力を活用したインタフェースの構築を図った。音声入力を活用した項目別自動入力インタフェースを 3 つ提案することにより、項目別の入力への音声入力の活用の活用性を示した。



図 4: 音声入力変更同期の例 (左: 音声入力文終端挿入による同期, 右: 音声入力文参照による同期)



図 5: ChatGPT に送信するプロンプトとレスポンス例

今後は看護師の協力のもと、3つの提案手法と手入力による電子カルテ入力の速度比較を行う予定である。また、本研究では音声認識部分に Web Speech API を使用したが、電子カルテ入力業務に特化した音声認識モデルを使用することにより、病院での実導入を想定したシ

ステムを構築する。

参考文献

- [1] Thomas G Poder, Jean-François Fiset, and Véronique Déry. Speech recognition for medical dictation: Overview in Quebec and systematic review. *Journal of Medical Systems*, Vol. 42, pp. 1–8, 2018.
- [2] Sherry Ruan, Jacob O Wobbrock, Kenny Liou, Andrew Ng, and James Landay. Speech is 3x faster than typing for english and mandarin text entry on mobile devices. *arXiv preprint arXiv:1608.07323*, 2016.
- [3] Julius Adorf. Web speech api. *KTH Royal Institute of Technology*, Vol. 1, , 2013.
- [4] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training. https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf, 2018.
- [5] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. *OpenAI blog*, Vol. 1, No. 8, p. 9, 2019.
- [6] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Matiusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Ilya Radford, Alec Sutskever, and Dario Amodei. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, Vol. 33, pp. 1877–1901, 2020.
- [7] Steve Krug. *Don't make me think!: a common sense approach to Web usability*. Pearson Education India, 2000.