聴覚障がい者向けの地声で対話するコミュニケーションツールの検討

越後 宏紀 * 金 韓栄 * 神田 悠杜 * 田中 航太 * 出戸端 賢人 * 山本 弘稀 *

概要. 既存の文字起こしツールや音声出力ツールにより現在では聴覚障がい者と健聴者が共にコミュニケーションを取りやすくなっている．しかしながら，現在はあくまで対話の中に翻訳ツールが仲介する構造になっており，聴覚障がい者と健聴者がシームレスに対話することは困難である．その問題を解決するために，我々は生成 AI を活用し入力される手話をリアルタイム翻訳しユーザの地声で音声出力するコミュニケーションツールの開発を目指している．本稿では開発の初期段階として，指文字をテキストに変換し，そのテキストをユーザの地声で音声出力するコミュニケーションツールを提案した．指文字をカメラで認識しテキストとして読み取り，そのテキストは生成 AI によって補完処理や文脈予測し対話の文を生成する．事前に 1 分弱の地声を登録しておくことで，テキストを地声で音声出力することを可能としている．今後は多言語手話への対応や，地声出力の効果を評価する実証実験を行う予定である．

1 はじめに

近年，障がいの有無に関わらず，誰もが安心して働ける環境を整備することが求められている．ソフトバンクでは，障がい者採用を実施している [5] 他，短時間での就業を可能とするショートタイムワークの推進など多様な人々が共に働く環境を整備している．

多様な人が共に働く環境を構築するために，健聴者と聴覚障がい者とのコミュニケーションの壁を減らすツールは数多く存在している．UD トーク [4] は，発話した声をリアルタイムで文字起こしを行い，テキスト出力でユーザに提示する．Zoom や Google meet などの Web 会議システムにも文字起こし機能が搭載されており，発話した声を認識し文字起こしを行い，テキスト出力でユーザに提示することが可能となっている．テキストを読み上げるツールも存在し，文字起こしされたテキストを事前に作成された多種多様なボイスで読み上げることが可能となっている．

我々は，健聴者と聴覚障がい者とのコミュニケーションの壁を減らすだけではなく，あたかも健聴者同士が自然に対話しているのと同等なコミュニケーションツールを目指して開発を進めている．本稿では，その初期開発として，ユーザの指文字をテキストに変換し，生成 AI を活用してテキストデータをユーザの地声で音声出力するツールを提案する (図 1).

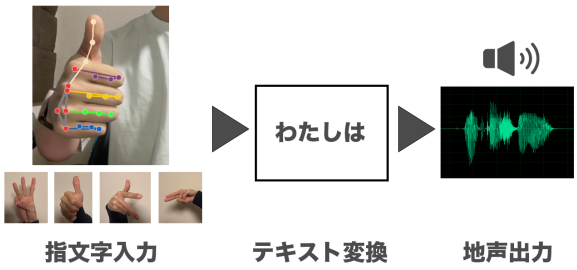

図 1. 本稿で提案するコミュニケーションツール

図 2. 提案ツールのインタフェース

2 指文字入力を地声で音声出力するコミュニケーションツール

2.1 提案ツールの基本構成

提案ツールのインタフェースは Web サイトで実装しており，HTML, CSS, Vue.js を活用して実装している (図 2). 入力は指文字、テキスト、音声入力が可能であり，入力されたデータは対話相手にテキストおよび音声で出力される．

* ソフトバンク株式会社

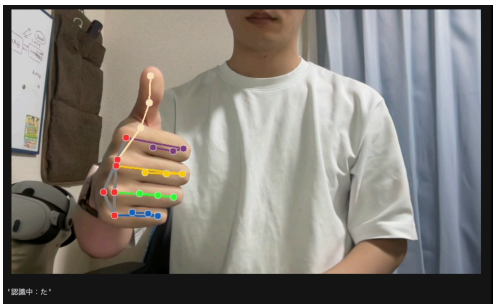

図 3. 指文字を認識している実装画面

2.2 指文字入力の実装

指文字入力の自動認識システムは先行研究 [9] が存在する．我々は Python で MediaPipe ライブラリを使用し，Web カメラで認識した指文字をテキストに変換するシステムを実装した．実際に指文字を検知しテキスト変換している様子を図 3 に示す．学習データは著者の指文字データを学習させ認識しているため，様々なユーザが使用する際の精度には制限がある．指文字や手話は言語同様多種多様であるため，それらのモデルの開発や精度向上については開発するのではなく，既存のモデルをアセットとして登録できるよう実装予定である．

2.3 テキスト変換の実装

Chat GPT-4o[1] を活用し，テキスト変換されたデータを自然な対話言語に変換する．このテキスト変換では，リアルタイムで入力されている間に発言したいことを予測する機能と，認識が欠如したデータを補完する機能を実装する．具体的には,「きょうのきぶんは」と入力されると,「今日の気分はどうですか？」といった文を予測し予測候補として出力される．また,「あし…のて…きは」と途中の入力が欠如していた場合,「あしたのてんきは」と文脈から予測補完し,「明日の天気はどうですか？」といった文を予測し予測候補として出力される．

2.4 地声を登録し音声出力する機能の実装

聴覚障がいは，人によって種類や程度が異なり，聞こえる聴力のデシベル (dB) や地声を発声できる感覚の有無も異なる．我々はまず，聴覚障がいの分類の主たる要素の 1 つであるアイデンティティに着目した．聴覚障がい者の分類の要素であるアイデンティティとは，先天性の障がいなどによって音声言語を習得する前に聴覚障がいになった「ろう者」，音声言語を習得した後，後天性の障がいで聴覚障がいとなった「中途失聴者」，障がいの発覚時期や程度には関わらず，手話や筆談ではなく残存する聴覚を使ったコミュニケーションを望む「難聴者」の 3 種が存在する [6]．中途失聴者や難聴者は，自身である程度言葉を発声することができるため，事前に音声を提案ツールに登録しておくことが可能である．本稿の提案するツールではろう者を対象から外しているが，ろう者が利用する場合はユーザの親戚によって登録する手法や，身体の体型，性別，年齢などの情報から推定した最適な音声を地声データとして登録可能とする予定である．本節では中途失聴者と難聴者を対象とし，提案ツールを用いて地声を登録する機能について記述する．

提案ツールでは，ユーザの発話レベルに応じてサンプル文を用意し，ユーザはそのサンプル文を発声することで地声を登録する．サンプル文は最大で 1 分程度であり，そのサンプル文のみのデータで様々な言葉に対応可能なボイスデータを作成する．発声された音声データは，Elevenlabs[2] により地声ボイスデータとして登録され，提案ツールでユーザが入力したデータを対話相手に地声のボイスとして出力することが可能とする．

3 今後の課題

3.1 多種多様な手話に対応する基盤の構築

日本では，日本手話，日本語対応手話，中間型手話の大きく分けて 3 種類の手話が存在する．世界では手話言語は 400 種類以上あると言われており，全ての手話データをモデルとして構築し対応することは困難である．我々は現在開発している指文字や手話データをベースとした上で，SureTalk[3] や各言語ごとに構築されたモデルを本提案ツールに登録し利用可能とするように実装していく予定である．

3.2 地声を含めた音声出力の印象評価と検証実験

地声や変声が対話者に対してどのような影響を与えるのか，見た目と声の印象のズレが対話者に対してどのような影響を与えるのかこれまで研究がされてきている [7][8]．本提案ツールではユーザである聴覚障がい者の地声を登録することで，ユーザの地声で音声出力し，より自然なコミュニケーションを実現できることを目指している．地声の出力によりコミュニケーションが従来より行いやすくなるのか，ユーザにどのような印象を与えるのかについて，今後実証実験を行い評価していきたいと考えている．

4 むすび

本稿では，健聴者と聴覚障がい者とのコミュニケーションをより自然に対話することを目指したコミュニケーションツールの初期段階で実装したツールを提案した．提案ツールでは，指文字入力からテキストに変換し，登録された地声で音声出力される．今後，多種多様な手話言語への対応を行うとともに，提案ツールが健聴者と聴覚障がい者とのコミュニケーションにどのような影響を与えるのか実証実験を行う予定である．

謝辞

本研究の開発でご協力くださったSureTalkの開発チームの皆様，および本提案ツールを開発するにあたって尽力いただいた岩崎響子さんに感謝を表する．

参考文献

[1] ChatGPT. https://openai.com/index/chatgpt/.

[2] Elevenlabs. https://elevenlabs.io/.

[3] SureTalk. https://www.suretalk.mb.softbank.jp/.

[4] UD トーク. https://udtalk.jp/.

[5] ソフトバンク障がい者採用. https://www.softbank.jp/recruit/disability/.

[6] 聴覚障がいとは？　等級や種類、コミュニケーション時に配慮すべきこと. https://www.suretalk.mb.softbank.jp/column/contents/000100.php.

[7] 呉健朗, 越後宏紀, 新井貴紘, 富永詩音. VTuberの変声と地声の違いによる印象評価. 情報処理学会シンポジウム論文集 (DICOMO2023), pp. 1643–1648, 2023.

[8] 呉健朗, 越後宏紀, 新井貴紘, 富永詩音. 異性アバタを用いるVTuberにおける変声の影響の検証. 情報処理学会シンポジウム論文集 (DICOMO2024), pp. 1205–1210, 2024.

[9] 渡邉聡, 五十嵐悠紀. Webカメラを用いた指文字自動認識システム. 第29回インタラクティブシステムとソフトウェアに関するワークショップ (WISS2021) 論文集, pp. 1–2, 2021.

未来ビジョン

本研究は，将来的にApple Vision Proを代表としたMRヘッドセットでの実装を予定している．MRヘッドセットで実装することで，シームレスに対話をすることが可能であると考えている．相手の手話を認識し，テキストで変換して提示することはもちろん，対話相手の地声で音声出力されることで，従来では実現できなかったより自然な対話を実現可能であると考える．

また，母国語が異なる話し手と聴き手が対話した際，翻訳者の声ではなく相手の地声で聴こえることで，より自然なコミュニケーションを実現できるのではないかと考える．他方で，地声ではなくユーザが理想とする音声での出力を望む場合，その音声を使用することでコミュニケーションにどのような影響を与えるのか検証を行う必要があると考える．

さらに，生成AIを活用し，場面や状況に応じた文脈を解釈しテキスト変換する機能も実装予定である．例えば，携帯電話ショップであれば携帯電話の購入や相談事項に最適な文を提示し予測変換する．場面や状況の解釈については，スマートフォンやPC，MRヘッドセットなどのデジタル機器からGPS情報を取得し，生成AIにその場面の情報も伝達することでより精度の高い予測変換が行えると考える．

我々は様々な障がいやコンプレックスを持っている人々が安心して暮らせる世の中を目指している．