

## 科学研究費助成事業 研究成果報告書

令和 5 年 6 月 9 日現在

機関番号：12103

研究種目：挑戦的研究(萌芽)

研究期間：2019～2022

課題番号：19K21745

研究課題名(和文)複数話者による同時多発音声の字幕提示手法に関する研究

研究課題名(英文) Research of the system displays dictated text of simultaneous plural voices by plural speakers

研究代表者

鈴木 拓弥 (SUZUKI, TAKUYA)

筑波技術大学・産業技術学部・教授

研究者番号：10553935

交付決定額(研究期間全体)：(直接経費) 3,800,000円

研究成果の概要(和文)：本研究は、複数話者によって音声と同時に発せられる場面において、これらの音声を字幕化し、視覚情報過多に陥ることなく受信できるような手法の開発を目的としており、予定通り実用可能なアプリケーションを開発し、公開した。研究の当初目的では聴覚障害者を主な受信者として定義していた。一方で開発した手法は即時共有性を重視しており、音声認識による字幕化以外に、文字入力を主としたコミュニケーションにおいても有効であることが分かった。十分な効果検証は今後の課題であるものの、開発した手法は現在も運用中であり、聴覚障害者、健聴者から授業、会議、各種のコミュニケーション目的など、幅広い用途で活用頂いている。

研究成果の学術的意義や社会的意義

本研究の成果であるアプリケーションは音声認識や翻訳等の機能を内包しており、幅広い活用が可能である。一方、即時共有性を重視したことで、当初重視していなかった文字入力についても、他のシステムにはない特性を有していることが分かった。複数音声を受信負荷の低い状態で聴覚障害者に伝達する目的で開発した手法である一方で、即時共有性は、聴覚障害者による発言、発信時に有効な手法であることが分かるなど、当初想定していなかった知見を得ることができた。現在も運用中であり、当事者意見を反映させ、アップデートを実施している。

研究成果の概要(英文)：The objective of this research was to develop a method to subtitle and receive speech without visual information overload in a situation where multiple speakers are speaking at the same time.

The initial objective of the study defined the hearing-impaired as the main receivers. On the other hand, the developed method emphasizes immediate sharing, and it was found to be effective not only in subtitling by speech recognition, but also in communication mainly through text input. The developed method is still in operation, and is widely used by both hearing-impaired and hearing-impaired people in classes, meetings, and various communication objectives.

研究分野：福祉情報工学

キーワード：音声認識 字幕 UI 支援 リアルタイム 同期 チャット 聴覚障害

## 1. 研究開始当初の背景

聴覚障害者に対する支援は様々実施されている一方、解決が困難な問題もある。例えば聴覚障害者の修学上、就業上の困難として、グループワークや会議における困難がある。聴覚障害者は複数話者が同時に話すような場面では、手話や字幕だけでは情報を十分に取得できない。そのため、資料の事前提示、複数話者が同時に話さない、司会進行が話者を調整するなどの配慮が行われている。

しかし、これらの配慮が常に行われるとは限らない。情報交換の質の減少と解釈されてしまう場合もある。この問題の解決には、これまでの支援や配慮とは基軸の異なる対策が必要である。ここでもし聴者が聴覚から情報を得るのと同等の情報取得機会を聴覚障害者が視覚から得ることができれば、本問題を緩和できるのではないかと。このように着想したことが本研究の背景である。

## 2. 研究の目的

本研究では、聴者を中心とした中に少数の聴覚障害者が参加するような会議やグループワークなど、複数話者によって音声と同時に多発的に発せられる場面において、これらの音声を漏れなく即時に字幕化して共有し、一方で視覚情報過多に陥ることなく、聴者が音声から得るのと等しい情報量を聴覚障害者が得られるような手法を開発し、検証することを目的とした。

## 3. 研究の方法

研究では、音声を認識し即時に文字化し、共有するためのアプリケーションを開発した。具体的には、話者それぞれに近接した PC やスマートフォン上で動作する音声認識アプリケーションと、それぞれのスマートフォンから送信されるテキスト情報を集約し、聴覚障害者の PC 若しくはスマートフォンに表示するための可視化アプリケーションを開発した。開発後、可視化システムに複数の字幕提示インタフェースを実装する。インタフェース実装後、聴覚障害当事者の協力を得て、複数話者による会話による内容理解度に関する実験と質問紙調査及びヒアリング調査を行った。

## 4. 研究成果

当初計画通り、Web アプリケーションを開発した。本アプリケーションはソーシャルアカウントの一つである Google アカウントを利用してログインして活用する。ログインユーザーは会話のための「ルーム」を設定でき、ユーザーはルームに参加し、会話する。図 1 はルームが 3 つ設定された状態を示しており、無限にルームを設置することができる。



図 1 ログイン直後の画面上部 (3つのルームが設定されている状態)

ルーム設置後、ユーザーは他の参加者を QR コードや招待コード、短縮コード等で招待する。参加者はログインせずにゲストすることが可能である。

図 2 はルームに 3 人が参加し、会話している様子を示している。

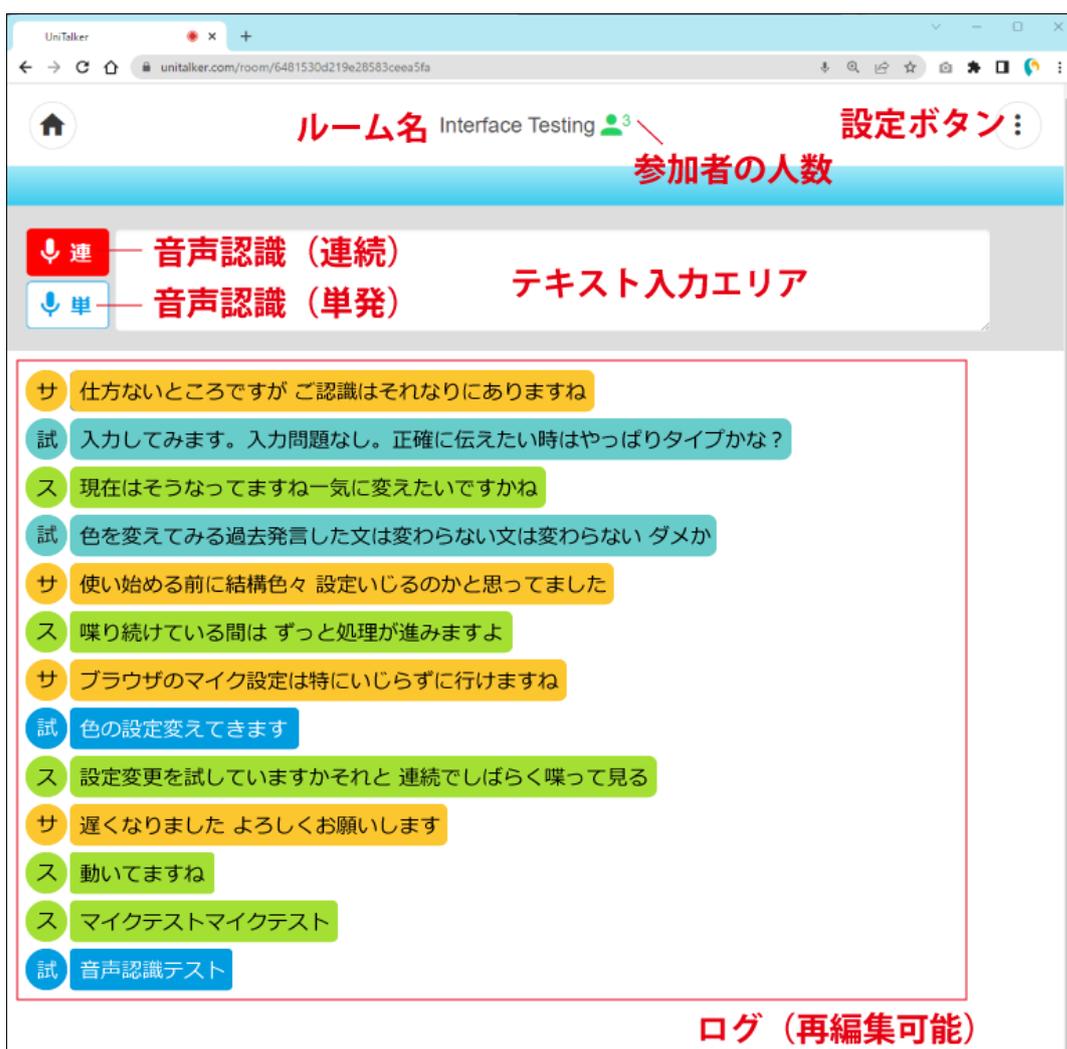


図 2 ルーム入室後の画面 (3 人で会話している様子)

「テキスト入力エリア」は文字入力及び音声認識結果の修正のために利用する。一般的なテキスト入力欄と同様である。右上の設定ボタンから設定できる時間の経過を待つ、あるいは画面右下に存在する決定ボタンを押して入力を確定する仕様である。「単」と表示されたマイクアイコンはトグル方式になっており、押すと音声認識が始まり、ボタンはグレー反転される。音声認識された結果は「テキスト入力エリア」に戻り、設定した時間だけウインドウ内に留まる。その間に音声認識結果を修正できるようになっている。決定ボタンを押すか、設定した時間経過によって入力を確定すると、変換後の文字が下部のメインウインドウに表示される。同時にグレー反転されていた「単」ボタンが元の状態に戻る仕組みである。つまり、発話毎に「単」ボタンを押す必要がある。「連」と表示されたマイクアイコンは音声入力毎にトグル処理とならず、ユーザーが「連」ボタンを再度押すまで、連続して音声認識と設定時間後のテキスト排出を繰り返す仕組みとなっている。

開発後、企業 3 社に依頼して試用する実証実験や、聴覚障害学生及び健聴学生 12 名に依頼した会話実験等を通じ、効果を検証した。

実験の結果、本アプリケーションの特性や有効性について明らかにすると共に、集団での会話に代表される双方向の情報伝達の必要な場面においては、発話と字幕提示のタイムラグを可能な限り無くすだけでなく、日本語特有のかな漢字変換前や入力確定前の変換過程も含めて遅延することなく即時に字幕を提示、共有することが、内容理解において重要である可能性が高いことが分かった。また、定性的評価により、音声をテキストデータに即時変換し、誤変換の修正が容易な UI とすることや、文字情報が一定量ログとして画面内に残ることで、聞き返しの必要がなく、聴覚障害者にとって精神的負荷の軽減につながるという評価も得られた。UI の改善提案を得ることもできた。

本ウェブアプリケーションは現在も運用中である。また、当事者意見を反映させ、アップデートを実施している。現時点で変換精度は十分とは言えないが、誤変換を推測、修正しやすい UI とすることで企業や障害当事者に活用頂いている。実験時同様、音声認識過程を含めて即時共有することで、誤変換が生じても正しい内容を推測しやすいという意見を頂いている。また、他者の補助を必要とせず、無償かつ簡単な手続きで導入や利用可能な点、ユーザーフレンドリーな点を評価頂いている。また、複数音声を受信負荷の低い状態で聴覚障害者に伝達する目的で開発した手法である一方、即時共有性は聴覚障害者による発言、発信時に有効な手法であることが分かるなど、当初想定していなかった知見を得ることができた。高い正確性が求められる場面では要約筆記やノートテイクなども併用しながら、障害当事者から寄せられる意見をフィードバックし、今後も活用と改良、運用を続けていきたいと考えている。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 Takuya Suzuki	4. 巻 2020PartIII
2. 論文標題 Simultaneous Speech Subtitling Systems for Multiple Speakers	5. 発行年 2020年
3. 雑誌名 HCI 2020. Communications in Computer and Information Science	6. 最初と最後の頁 114-120
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-50732-9_16	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Takuya Suzuki, Ryoga Murate, Masafumi Nishida 1580	4. 巻 1580
2. 論文標題 Development and Evaluation of UniTalker, an Application for Simultaneous Presentation of Subtitles from Multiple Speakers	5. 発行年 2022年
3. 雑誌名 HCI 2022. Communications in Computer and Information Science	6. 最初と最後の頁 612-619
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-031-06417-3_82	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計1件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 鈴木拓弥
2. 発表標題 複数音声同時字幕提示ユーザインタフェースの開発と効果検証
3. 学会等名 情報処理学会アクセシビリティ研究会第14回研究会
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------