

聴覚障害者支援のためのマルチモーダル発話可視化に関する研究

鳥羽 祐輔[†] 堀内 大祥[†] 杢本 真佑[†] 佐伯 幸郎[†] 中村 匡秀[†]
内野 智仁^{††} 横山 知弘^{††} 武林 靖浩^{††}

[†] 神戸大学 〒 657-8501 神戸市灘区六甲台町 1-1

^{††} 筑波大学附属聴覚特別支援学校 〒 272-8560 千葉県市川市国府台 2-2-1

E-mail: [†]toba@ws.cs.kobe-u.ac.jp, ^{††}masa-n@cs.kobe-u.ac.jp

あらまし 聴覚障害者のコミュニケーション手段としては手話や筆談, 口話法など様々な方法が存在するが, 会議などの複数人かつ対面という状況では, それぞれ欠点を持つ. 本研究では, 聴覚障害者に対する健聴者の発話内容の理解支援のためのマルチモーダル発話可視化アプリケーションを提案する. 本研究におけるマルチモーダルな可視化とは, 複数の情報の提示・可視化モードを組み合わせた複合的な情報の提示手段のことを指す. 音声認識を使った発話内容のテキスト化モードや, 専門用語の説明モードなどを利用者自身が選択することで, 様々な聴覚障害のレベルやニーズに合わせた情報保障を実現できると考える. さらに予備実験として, Google が提供しているオープンな音声認識エンジン, Web Speech API の発話認識精度の評価実験を行う.

キーワード マルチモーダル発話可視化, 聴覚障害者, コミュニケーション支援, 口話法, 音声認識, Web Speech API

A Study of Multi-Modal Speech Visualization for Deaf and Hard of Hearing People Support

Yusuke TOBA[†], Hiroyasu HORIUCHI[†], Shinsuke MATSUMOTO[†], Sachio SAIKI[†], Masahide
NAKAMURA[†], Tomohito UCHINO^{††}, Tomohiro YOKOYAMA^{††}, and Yasuhiro
TAKEBAYASHI^{††}

[†] Kobe University, Rokkodai 1-1, Nada, Kobe, Hyogo, 657-8501 Japan

^{††} Special Needs Education School for the Deaf, University of Tsukuba,
2-2-1 Konodai, Ichikawa City, Chiba, 272-8560 Japan

E-mail: [†]toba@ws.cs.kobe-u.ac.jp, ^{††}masa-n@cs.kobe-u.ac.jp

Abstract Although deaf and hard of hearing (D/HH) people have various communication ways such as sign language, conversation by writing, and lip-reading, each way has several disadvantages in one-to-many conversation. This paper proposes a multi-modal visualization application in order to support D/HH people in understanding conversation with hearing people. Here, the term *multi-modal visualization* means providing informative contents by presenting multiple visualization modes. We suppose that selecting and combining each visualization mode by D/HH people themselves would be helpful for understanding of conversation. Additionally, we conduct a preliminary experiment for evaluating recognition accuracy of Web Speech API, which is voice recognition open software provided by Google.

Key words Multi-modal speech visualization, deaf and hard of hearing people, supporting understanding conversation, lip-reading, voice recognition, Web Speech API

1. はじめに

先天的,あるいは後天的に聴覚に障害を持つ聴覚障害者に対

し,手話や字幕等の代替手段を用いて情報を提供することを情報保障と呼ぶ.大学等の教育現場では,情報保障の一環として,

一般学生が講義の内容を書き出し伝えるノートテイク（要約筆記）のボランティアを設けていることも多い。また、平成 25 年には障害者雇用促進法が改正されており、障害者の雇用や就労などの社会進出は年々着実に進展しつつある。そのため、情報保障は教育現場や自治体のみならず、社会全体で取り組むべき重要な事項となっている [1][2][3][4]。

しかしながら、聴覚障害は外見上判断が付きにくいという特性があり、周辺環境の理解や情報保障の状況は十分とはいえない [5][6][7]。職場において会議や研修などの内容が十分に理解できないため、昇進の機会を失ってしまうといったケースも報告されている [5]。聴覚障害者の福祉向上のためには、会議などのような健聴者の集団内において、聴覚障害者が正しく情報を取得し発信できるような情報保障の枠組みが不可欠である [3][4]。

聴覚障害者のコミュニケーション手段には様々なものがある。その多くは障害のある聴覚の能力を視覚能力で補う手段である。最もよく知られている方法は手話であり、手や指、腕、表情などを用いた視覚言語によりコミュニケーションを取る。この手話は通常の会話と同等の自然な速度でコミュニケーションが可能という利点を持つ。他にも文字を用いる筆談や、発話者の口の動きを読んで内容を読み取る口話法などが使用される。

これらの手段は一對一の対面という状況では有効であるが、会議などの複数人かつ対面という状況では、それぞれ欠点を持つ。まず、手話は自由に使いこなせるようになるには学習コストが高い。また、聴覚障害者の中でも手話ができる割合は 14.1%という調査結果 [8] もあり、その普及率は決して高くない。筆談は識字率が 100%近い日本においては有効な手段ではあるが、その性質上複数人でのコミュニケーションには適さない。口話法は口の動きを読み取るという特性上、発話者全員と常に正対できないような状況では有効に働かない。

本研究では、情報保障手段の一つとして、聴覚障害者に対する健聴者の発話内容の理解支援という課題に取り組む。特に、会議等の複数人かつリアルタイムな対面コミュニケーションという状況を想定する。この課題を解決するために、発話内容に関する様々な情報を切り取り、効果的に表示するマルチモーダル発話可視化アプリケーションを提案する。ここでのマルチモーダルな可視化とは、複数の情報の提示・可視化モードを組み合わせた複合的な情報の提示手段のことを指す。モードの一例としては、マイクデバイスと音声認識エンジンを利用して、発話内容をテキスト化して可視化する STT (Speech to Text) モードが考えられる。他にも、カメラデバイスと顔認識エンジンを利用し、話者の口元を拡大化して表示する口話法補助モードなどもその一つである。これらの可視化モードを、聴覚障害者自身が選択し組み合わせ相互補完的に利用することで、様々な聴覚障害のレベルやニーズに合わせた情報保障を実現できると考える。

本稿では上記のマルチモーダル発話可視化アプリケーションの提案に加え、予備実験として Web Speech API^(注1)の発話

認識精度の評価実験を行う。これまでに、聴覚障害者のための STT システムはいくつか提案されている [9][10] が、その多くは有償の音声認識ソフトウェアを用いたものであった。一方で、現在では音声認識ソフトウェアはクラウドサービスという形で Web 上に一般公開されており、極めて低コストで音声認識処理の導入が可能になりつつある。このようなオープンなサービスが、どの程度の発話認識精度を確保できるか、情報保障システムに適用可能かといった点について実験、考察を行う。

2. 準備

2.1 聴覚障害者のコミュニケーション手段

聴覚障害者が日常生活において使用するコミュニケーション手段としては以下のようなものがある。いずれの手段にも長所と短所があり、実際はこれらの手段を組み合わせるなどして短所を補い合うような形で使用することが多い。

2.1.1 手話

手や指、腕、表情などを用いた視覚言語によりコミュニケーションを取る手法である。音声言語と同程度の自然なスピードでコミュニケーションできるといった利点を持つ。しかし自由にコミュニケーションができる程度の手話の修得には、高い学習コストが必要である。また、因果関係や抽象概念を表現能力が低く、コミュニケーション効率が低いという指摘もある。

2.1.2 筆談

文字を用いてコミュニケーションを行う手法であり、近年ではスマートフォンなどの IT 端末を用いた筆談アプリも数多く提供されている。識字率が 100%近い日本においては筆談は有効な手段であり、特殊な技術が不要という特徴もある。また手話と比べて伝達できる情報の正確性が極めて高く、約束事や重要な点を伝えるような状況に適する。一方でコミュニケーション速度や効率に課題があり、通常の会話と比べて多くの労力と時間を要する。そのため、会議のような場面では参加者全員の理解と協力が必要となるという難しさも存在する。

2.1.3 口話法

話者の口の動きから発話の内容を読み取る方法のことであり、聴覚口話法や読話とも呼ばれる。聾学校で学んだり、自然と口型を見る習慣を身につけることで修得できる技術である。健聴者はある程度はつきりと口を動かすように意識すればよく、特別な技術や道具が不要という点で健聴者に対する導入コストは低い。ただし、この方法は補聴器から聞こえる情報を参考にしつつ、前後の文脈から推測・補完するものであり、口話法単独で全ての情報が伝わる訳ではない。さらに、話し相手と正対し口が見える状態でなければ使用が難しく、また専門用語や固有名詞など知らない単語は推測することも困難である。

2.2 研究のスコープと解決すべき課題

本研究では、聴覚障害者のための情報保障手段の一つとして、健聴者の発話内容の理解支援という課題に取り組む。特に、会議などのような健聴者集団の中でのリアルタイムな対面のコミュニケーション状況を想定する。2.1 節で述べたコミュニケーション手段は、組み合わせることでそれぞれの短所を

(注1): <https://www.google.com/intl/ja/chrome/demos/speech.html>

る程度補い合うことができるが、本稿で想定する複数人での状況への対応は難しい。会議出席者の全員が手話を使えるというような状況は実現が難しく、また議論の場における筆談の利用は会議参加者の強い理解と協力が必要となる上、コミュニケーション効率の低下が避けられない。さらにそのような場面では全ての発話者と常に対面関係にあるとは限らず、口話法の使用も難しい。

このような情報保障手段の不足により、多くの聴覚障害者は会議への出席そのものが困難になっており、昇進の機会を失っている。情報保障制度が整っているイギリスなどの国では聴覚障害者は会議にも参加でき、能力さえあれば幹部などのポジションに任命されることも珍しくないことを考慮すると、会議へ参加できないことは働く聴覚障害者にとって非常に大きな障害となっている。聴覚障害者の福祉向上のためには、会議などのような健聴者の集団内における情報保障手段の提供が重要となる。

3. 聴覚障害者向け会話内容理解支援アプリケーション

3.1 機能要件

本アプリケーションは聴覚障害者にとっての健聴者の発話内容の理解度を向上させることを目的とする。そのために以下の要求を満たすようなアプリケーションを提案する。以下の要求は、筑波大学附属聴覚特別支援学校に所属する著者らの経験から洗い出されたものである。

R1 低導入コスト：まず第一に、健聴者と聴覚障害者の両者に対する導入コストをいかに押さえられるかが重要となる。ここでの導入コストとは、情報保障システムの環境構築に必要な特殊なデバイスやソフトウェアのほか、健聴者・聴覚障害者の両者の学習コストも含まれる。大学や企業など様々な状況での普及を目指す場合、高価なデバイスやソフトウェアの利用は避けるべきである。可能な限りスマートフォン等の身近なデバイスや、クラウド上に公開されたオープンなサービスのみでシステムが実現されることが望ましい。

R2 発話内容理解に役立つ情報の提示：発話内容を理解する際は様々な観点から情報を取得することができる。手話であれば手の動き、筆談であれば文字、口話法であれば口型などがそれにあたる。発話内容を理解するのに必要とされる様々な情報を取得し、それらを組み合わせて聴覚障害者に分かりやすいように視覚的に情報提供することで発話内容の理解を支援することが必要である。例えば、発話内容を音声認識処理しテキスト化して提示する方法や、発話者の口型の映像を表示するような機能などが該当する。

R3 発話に関する情報の振り返り：実際の会議等のシチュエーションでは複数人が同時に発話することもあり、リアルタイムに口型から発話を読む口話法では対応しきれない。したがって、必要に応じて過去の発話に関する情報を参照できる機能は発話内容の理解に貢献できると考えられる。

また議事録を自動生成したいという要求もある。従来の情報

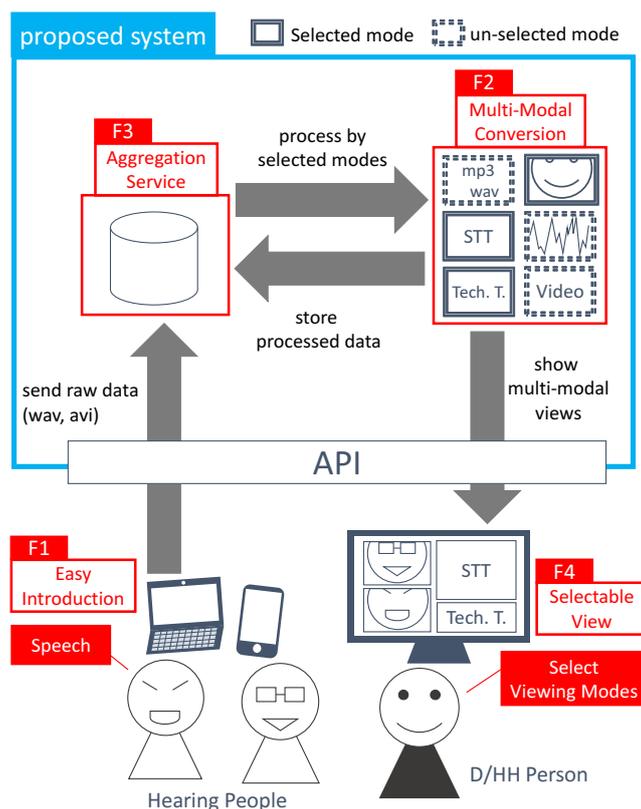


図 1: 提案アプリケーションのアーキテクチャ

保障は手話通訳やパソコン要約速記など人の手を介するものが主流でありそのコストは大きく、そのコストの大きさが情報保障を難しくしている要因となっているからである。議事録が自動生成され発話に関する過去の情報を振り返ることができれば情報保障に貢献できると考えられる。

R4 聴覚障害者自身による情報の取捨選択：要件 R2 では様々な情報提示機能が必要であると定めたが、同時に数多くの情報が提示されることで情報過多となり、理解の妨げとなってしまうケースも存在する。例えば、テレビ視聴時に質の悪い、あるいはリアルタイム性が欠けていて先に結論が分かってしまうような字幕を、あえて非表示にするといったケースもある。また、発話内容理解の補足となる情報は、聴覚障害者自身の聴覚障害の程度や、手話のスキル、口話法の修得度合いに強く依存する。これらの点から、複数の情報提示・可視化機能を聴覚障害者自身がその必要性に応じて選択できることも重要な要件となる。

3.2 アーキテクチャ

提案アプリケーションのアーキテクチャを図 1 に示す。利用状況の前提として、会議での各発話者がスマートフォンやノート PC などの端末を用意する。これらの端末に組み込まれたマイクとカメラを通じて、健聴者の発話情報となる音声データや画像データがシステムのサーバ内に蓄積される。提案アプリケーションは Web アプリケーションとして実現されるため、身近に存在するスマートフォンやノート PC など様々な端末から利用できる。蓄積された音声と画像データは、聴覚障害者が選択した各種のモードにより処理され、ユーザの所有するデバイス上に表示される。例えば STT (Speech to Text) モード

を用いれば発話内容をテキストベースに可視化したものを表示し、Tech. T.(Technical Term) モードを用いれば発話された専門用語や固有名詞などに関する補足情報を表示することができる。また、各モードによって変換されたデータはシステムのサーバ内に蓄積され、元の音声データや画像データと共に必要なときに利用可能である。

3.3 機能

本節では提案アプリケーションの利用シーンの例に基づいて4つの機能を説明する。

F1 様々な端末で利用可能：提案アプリケーションは Web アプリケーションとして実装される。したがってスマートフォンや PC 上のブラウザを使うことによってあらゆる場所から特別なデバイスやソフトを購入することなく利用することができる。近年のスマートフォンや PC の所有率を考慮するとこの機能によって多くの健聴者や聴覚障害者が提案アプリケーションを容易に使用できると言える。想定される利用シーンを図2に示す。

F2 会話理解支援情報の取得と提示：健聴者の発話内容を理解する手助けとなるような情報を取得し、ユーザに提示する。提供するモードは自由に追加することができる。本稿では以下3つの情報提示モードから発話内容の理解を支援するアプリケーションを提案する。本機能の表示例を図3に示す。

- 発話内容テキスト化モード：健聴者の発話内容を音声認識処理を通してテキストベースに可視化する。単なる発話内容だけでなくその発話者や会議情報をセットにして取得することも可能である。本機能の表示例を図3の右上に示す。

- 発話者の口型表示モード：口話法の補助として発話者の口元の映像を表示する。本機能は正面から向き合っていない発話者に対しても発話者が所有するスマートフォンや PC に搭載されたカメラを通してリアルタイムに口元の映像を表示することで口話法を補助するものである。健聴者も発話時に体の向きを意識せず自然に発話できるようになり、より自然なコミュニケーションの中で口話法が行えるようになる。本機能の表示例を図3の左側に示す。

- 類推困難な単語の説明モード：専門用語・固有名詞など口話法での読み取りが困難な単語をテキスト情報から抽出し、その詳細情報と共に表示できる。口話法では聴覚障害者は話の内容や流れを読みながら言葉を推測・予測するので専門用語や固有名詞など知らない単語は読み取ることが困難である。この機能によって聴覚障害者は口話法では読み取ることのできなかつた言葉の理解を文字という視覚情報で補助することができる。本機能の表示例を図3の右下に示す。

F3 サーバに発話情報をためる：システム内のサーバに発話情報を蓄積することで過去の発話内容を振り返ることができる。例えば会議の最中に発話の一部分を理解しそこねた場合やリアルタイムには理解できなかった場合に数秒前の会話情報を振り返り、テキスト化されたものや発話者の口型を確認することで会話内容を理解することができる。

また本機能は簡易議事録自動生成機能としても活用できる。

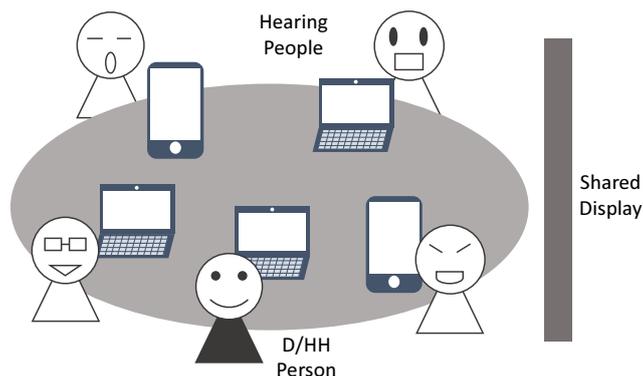


図2: 提案アプリケーションの使用シーン例

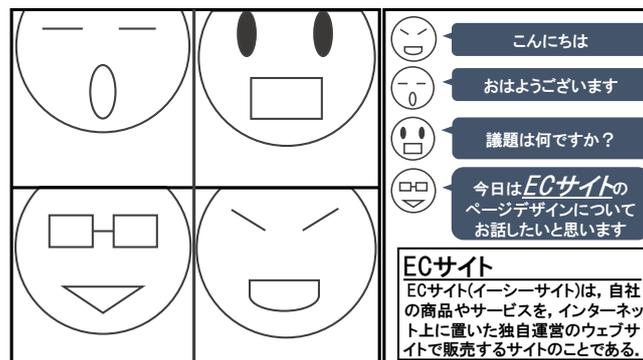


図3: 提案アプリケーションのディスプレイ結果例

聴覚障害者への情報保障の観点から見ても人の手を介さずとも容易に議事録が生成できることは、聴覚障害者自身にとっても雇用する企業側にとっても有益なものである。

F4 ユーザが取得する情報の種類を選択できる：F2で提案した情報提示モードの中からどの情報を使用するかはユーザ自身が選択できる。またその情報を表示する画面構成もユーザが使いやすいように配置できる。例えば図2、図3のように実際の会議での席の配置とディスプレイ上に口元情報を表示する配置を対応づけることでアプリケーション上での情報を分かりやすく整理することもできる。発話内容を理解するために必要な情報をそれぞれの会話シーンに応じてユーザ自身が調整していくことができるので、効果的に発話内容理解を支援することができる。例えば少人数の会議で口元映像を取得せずとも口型を見ることができると判断した場合はユーザはテキスト情報と用語解説の機能だけを使用することもできる。

4. 実装

本節では本アプリケーションのプロトタイプの実装について説明する。実装環境は以下の通りである。

- 開発言語：Java, JavaScript, HTML, CSS
- 音声認識エンジン：Web Speech API

今回は提案した機能の中で以下にあげる一部の機能を実装したプロトタイプを開発した。

発話内容テキスト化モード：ボタンを押すと音声認識モードになり、発話者のデバイスに付属するマイクを通して音声デー

タを取得する．そのデータを Google の Web Speech API を用いてテキスト化する．そして WebSocket を通じてそのテキスト情報と発話者情報をセットにして全てのアプリケーション使用者のブラウザ上に送信・共有する．聴覚障害者側のアプリケーションのテキストログエリアに表示され、自分がそのアプリケーションを起動してから現在までのテキストログはスクロールラブルに参照できる．

発話者の口型表示モード：発話者の映像は Web カメラを通して取得され、WebRTC を用いて全ての発話者のブラウザから聴覚障害者のデバイスへリアルタイムに送信される．聴覚障害者はアプリケーションを通して全ての発話者の映像を確認できる．発話者は自分の映像だけを確認できる．

類推困難な単語の説明モード：発話内容テキスト化モードによって表示された文章の中で、マウスのドラッグ操作により選択された単語の意味を表示することができる．選択された単語を JavaScript のプログラム内で取得し、その情報を MediaWiki API に通すことによって Wikipedia にその単語のページが存在する場合は冒頭の説明文を取得・表示できる．

5. Web Speech API の発話認識精度実験

5.1 実験概要

本節では、提案アプリケーションの STT (Speech to Text) エンジンとして用いた Web Speech API の発話認識精度に関する予備実験について説明する．これまでに、有償の STT エンジンを用いた聴覚障害支援方法が提案されているが、現在では無償の STT サービスも増えつつある．このようなオープンな STT サービスによる、聴覚障害者支援システムへの適用可能性について検討する．

提案システムの利用状況としては様々な場面が考えられる．実験では、3 種類の環境属性 (マイクの種類、周辺雑音の強さ、発話文章の種類) をそれぞれ組み合わせることで、様々な環境下での認識精度を確かめる．

5.2 実験設定

実験で制御する 3 種類の環境属性を表 1 に示す．マイクデバイス (表 1a) としてはノート PC, スマートフォン, 及びヘッドセットマイクを用いる．スマートフォンについては卓上に設置した場合と口元に近づけた場合の 2 種類を考える．外部環境 (表 1b) としては、静かな環境と騒がしい環境の 2 種類を用意する．騒がしい環境は人の多いカフェの自然音を録音したものを大音量で流すことで再現する．発話文章 (表 1c) は日常的に用いられる 3 つの文章 ($sent_1$ - $sent_3$) と専門用語を多く含む 2 つ文章 ($sent_4$ - $sent_5$) を用いる．実験ではこれら属性の全ての組み合わせ 40 通り (4 マイク \times 2 外部環境 \times 5 発話文章) について、それぞれ 10 回の発話を繰り返し、その平均認識精度を測定する．

5.3 評価尺度

認識精度の評価尺度として、音素ごとの認識の正しさを計測する．式 (1) に実験に用いた認識精度を示す．

$$\text{認識精度 (\%)} = \left\{ 1 - \frac{\text{認識誤り音素数}}{\text{発話文章の総音素数}} \right\} \times 100 \quad (1)$$

表 1: 実験で制御する 3 種類の環境属性

(a) マイクデバイス		(b) 外部環境	
ラベル	概要	ラベル	概要
$mic_{headset}$	ヘッドセットマイク	env_{quiet}	静か
$mic_{sp:mouth}$	スマートフォン (口元)	env_{noisy}	騒がしい
$mic_{sp:desk}$	スマートフォン (卓上)		
mic_{laptop}	ノート PC 付属のマイク		

(c) 発話文章	
ラベル	概要
$sent_1$	こんにちは
$sent_2$	私はその意見に賛成です
$sent_3$	それでは前回の会議のおさらいから始めたいと思います
$sent_4$	クラウドコンピューティングとは様々な計算資源をネットワーク上のサービスとして提供・利用する新しいコンピューティングパラダイムです
$sent_5$	クラウドコンピューティングのサービスモデルには IaaS, PaaS, SaaS の 3 つの形態があります

表 2: 発話認識の精度 (%)

(a) 静かな環境 (env_{quiet})					
	$sent_1$	$sent_2$	$sent_3$	$sent_4$	$sent_5$
$mic_{headset}$	100.0	95.0	95.3	96.5	93.1
$mic_{sp:mouth}$	100.0	100.0	97.3	97.8	94.2
$mic_{sp:desk}$	100.0	100.0	95.7	95.9	92.9
mic_{laptop}	100.0	88.1	75.3	95.1	82.3

(b) 騒がしい環境 (env_{noisy})					
	$sent_1$	$sent_2$	$sent_3$	$sent_4$	$sent_5$
$mic_{headset}$	100.0	100.0	90.3	83.9	81.0
$mic_{sp:mouth}$	100.0	99.4	97.3	90.1	91.4
$mic_{sp:desk}$	100.0	81.3	73.3	88.7	86.9
mic_{laptop}	100.0	61.3	53.7	65.1	62.5

この式では、例えば「今日は晴れ」という文章が「今日はあれ」と認識された場合、「ハ」という音素が「ア」と誤認識されているため、総音素数 5 つに対する認識精度は 5 分の 4、すなわち 80.0%となる．

なお、Web Speech API では発話音素の認識処理のみならず、認識結果の漢字変換処理も同時に行われる．そのため「今日は腫れ」のように、音素は正しく認識されていても漢字変換の処理誤りが発生する可能性がある．定量的な認識精度の結果としては、上記の通り音素レベルでの認識精度のみに着目し、この漢字変換の誤りは無視するものとする．

5.4 結果と考察

表 2 に、40 通りの全ての組み合わせの認識実験結果を示す．実験結果より、発話認識に対する理想的な状況、すなわち静かな環境 (env_{quiet}) で、マイクとの距離が近く ($mic_{sp:mouth}$)、発話者が意識して明確に発話するような環境であれば、どのような文章でも 90%以上の認識精度を確保できている．しかしながら、騒がしい環境 (env_{noisy}) かつノート PC 付属のマイク

表 3: 実験で観測された認識誤りの一例

例 1	正	それでは前回の会議のおさらいから始めたいと思います
	誤	それでは前回の会議のお皿から始めたいと思います
例 2	正	私はその意見に賛成です
	誤	私はこの意見に賛成です
例 3	正	クラウドコンピューティングとは様々な計算資源をネットワーク上のサービスとして提供, 利用する新しいコンピューティングパラダイムです
	誤	クラウドコンピューティングとは 様々な計算資源をネットワーク嬢のサービスとして提供利用する新しいコンピューティングパラダイムです

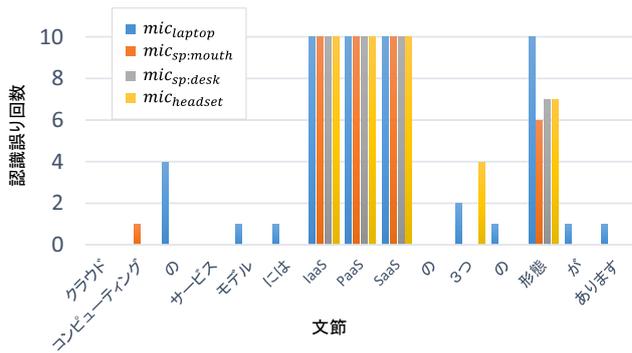


図 4: 文章 *sent5* における単語毎の認識誤り回数

(*mic_{laptop}*) では, その認識精度はおよそ 50% から 60% 程度と低かった. 会議等の複数人の場面を想定した場合, このようなノイズの多い環境は容易に発生すると考えられる. これらの結果より, 理想的な環境ならば Web Speech API は提案システムの STT エンジンとして適用可能ではあるが, より現実に即した環境では何らかの工夫が必要であるといえる.

次に, 具体的な認識誤りの結果について述べる. 実験で観測された認識誤りの一例を表 3 に示す. 表に示すように, 音素そのものの認識誤り (例 1 と例 2) だけでなく, 漢字変換の誤り (例 3) も多く存在していた. 著者らの聾学校での経験から, 漢字間違いが含まれる文章は, 全てひらがなの文章よりも理解が困難であるという意見も得られている. したがって, モードの一つとして漢字変換を行わず, 認識された音素のみをひらがなだけで表示すると言った機能も効果的であると考えられる.

文章中での具体的な誤りの傾向を確認するために, 図 4 に文章 *sent5* の発話の際に発生した単語ごとの認識誤り回数を示す. この図は静かな環境における, 各マイクデバイスごとの認識誤りを表している. なお縦軸の認識誤り回数とは, 音素誤りと漢字誤りの両方を合算したものである. 結果より, IaaS/PaaS/SaaS などの専門性の高い用語は, 10 回の試行の内, 一度も正しく認識されなかった. 具体的には, 「IaaS」の場合「愛アース」や「ir」などと認識されていた. この原因は, 認識処理に用いられる辞書の中に, その単語が含まれていなかったことにあると考えられる. 会議の内容やテーマに合わせたドメイン辞書を, 認識処理に用いるといった対策が必要であるといえる.

6. おわりに

本稿では, 情報保障手段の一つとして, 発話内容に関する様々な情報を切り取り, 効果的に表示するマルチモーダル発話可視化アプリケーションを提案した. さらに提案システムのプロトタイプを開発し, オープンな STT エンジンの発話認識精度に関する評価実験を行った.

現在, 提案システムはプロトタイプとして一部の機能のみが実装されており, システムの作り込みと不足機能の充足, 及びシステム全体の被験者実験は今後の重要な課題である. また, システム単体で全ての問題を解決するのではなく, 人手による支援そのものを助ける仕組みを検討中である. 音素の似通った指示語の認識誤りなどは, 音声認識技術のみで解決することは困難である. この認識誤りを人手によって修正できる仕組みと, その修正作業をうまく動機付けさせられる仕組み (ゲーミフィケーション等) を用意することで, システム単体での能力不足を人手で補える可能性が高い. さらには, その修正作業を通じて聴覚障害そのものに対する理解を支援できるようなアプリケーションの開発に取り組んでいきたい.

謝辞 この研究の一部は, 科学技術研究費 (基盤研究 C 24500079, 基盤研究 B 26280115, 基盤研究 C 24500258, 若手研究 B 26730155), および, 川西記念新明和教育財団の研究助成を受けて行われている.

文 献

- [1] M. Marschark, G. Leigh, P. Sapere, D. Burnham, C. Convertino, M. Stinson, H. Knoors, M.P. Vervloed, and W. Noble, "Benefits of sign language interpreting and text alternatives for deaf students' classroom learning," J. Deaf Studies and Deaf Education, vol.11, no.4, pp.421-437, 2006.
- [2] A.M. Piper and J.D. Hollan, "Supporting medical conversations between deaf and hearing individuals with tabletop displays," Conf. Computer Supported Cooperative Work, pp.147-156, 2008.
- [3] R. Punch, P.A. Creed, and M.B. Hyde, "Career barriers perceived by hard-of-hearing adolescents: Implications for practice from a mixed-methods study," J. Deaf Studies and Deaf Education, vol.11, no.2, pp.224-237, 2006.
- [4] R. Punch, M. Hyde, and D. Power, "Career and workplace experiences of australian university graduates who are deaf or hard of hearing," J. Deaf Studies and Deaf Education, vol.12, no.4, pp.504-517, 2007.
- [5] 水野映子, "聴覚障害者が働く職場でのコミュニケーションの問題-聴覚障害者・健聴者に対するアンケート調査をもとに-", Life Design Report, 2014.
- [6] 一般財団法人全日本ろうあ連盟, "聴覚障害者の福祉施策に関する要望について," <http://www.jfd.or.jp/2015/01/21/pid13028>, 2010.
- [7] A. Weisel and R.G. Cinamon, "Hearing, deaf, and hard-of-hearing israeli adolescents' evaluations of deaf men and deaf women's occupational competence," J. Deaf Studies and Deaf Education, vol.10, no.4, pp.376-389, 2005.
- [8] 厚生省大臣官房障害保健福祉部, "身体障害者・児実態調査結果の概要," http://www1.mhlw.go.jp/toukei/h8sinsyou_9/, 1999.
- [9] 服部裕之, "音声同時字幕システム," 第 3 回 FIT 講演論文集 CD-ROM, 2004.
- [10] 磯野春雄, 山口淳平, "音声認識技術を利用した聴覚障害学生のための字幕講義システム," 日本人間工学会大会講演集, vol.45, no.0, pp.204-205, 2009.