

講義情報保障における音声認識の活用事例

～経験に基づき作成した情報保障アプリの紹介～

西俣 稔子

本校高等部専攻科 歯科技工科は 1971 年の設置以来、歯科技工士養成機関として専門的教育を行ってきた。専門分野の講義は近隣の大学歯学部的外部講師によるものが多く、その場合の情報保障は内部教員が担ってきた。2005 年に情報保障に音声認識を試験的に導入して以来、試行錯誤を重ねてきたが、現在は自作した情報保障用 Web アプリを利用して、精度の高い情報保障が少ない労力でできるようになった。本稿ではこれまでの経緯を報告し、本アプリの効果を検証し、情報保障における音声認識の有用性を考察する。

キー・ワード：情報保障 音声認識 専門用語 Microsoft Edge Google Chrome

1 音声認識の導入と実践

歯科技工科で最初に利用した音声認識は「ドラゴンスピーチ」というパッケージソフトで、2005 年に「セレクト 7」、2012 年に「11」を導入した。当時は音声認識に適した独特な話し方をしないと、認識されなかったもので、講師の話し方では音声認識に適さず、復唱する必要があった。したがって、情報保障には復唱者と修正者の少なくとも 2 名以上、PC も 2 台以上必要だった。

その後、スマートフォンが普及し始めた頃から、サーバー型の音声認識の精度が飛躍的に向上し、UD トークや Google 音声文字変換などの情報保障アプリが出され、夢だった講師の音声直接入力現実的になった。特に Google の音声認識は語彙が豊富で歯科専門用語にも比較的対応できていた。ただし、長めの専門用語は認識するが、短めの専門用語はどうしても同音異義語に誤変換されてしまう。ドラゴンスピーチは辞書登録のカスタマイズでそのような専門用語も出るようにできたが、サーバー型では難しくなった。

2017 年に IPtalk に追加された s2t_Chrome ウィンドウ機能により、Google 音声認識の利用と誤変換の修正が可能になったので、講師の音声直接入力、修正者 1 名、PC 1 台での情報保障に取り組んだ。認識精度は良かったが、ウィンドウが多く画面が複雑になってしまい、操作も煩雑だった。また、歯科技工では、「ここ」などの指示語がよく使われるが、表

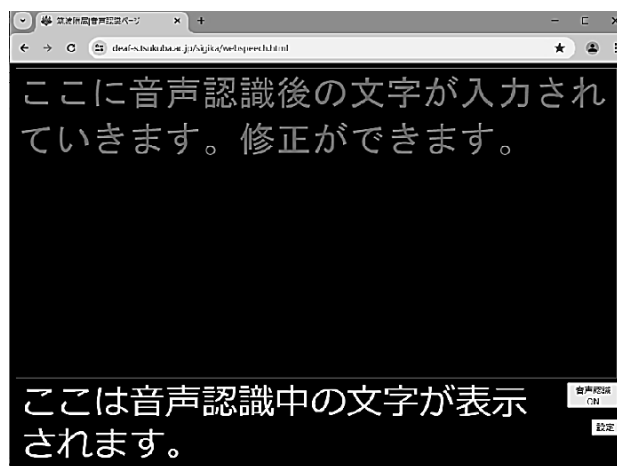
示が遅れると何処を指すのか分からなくなってしまうので、音声認識中の文字もリアルタイムで表示させたいが、IPtalk_s2t_Chrome ウィンドウはそうには作られていなかった。

そこで、画面がシンプルでリアルタイム表示もできる Google ドキュメントの利用を考えたが、修正箇所カーソルを移動させた時に音声入力があると、その位置に音声認識された文字が入ってきてしまう問題（以下、「カーソル問題」）があり、実用には至らなかった。

2 アプリの自作

これまでの実践経験から、求める機能は、Google 音声認識でリアルタイム表示ができて、カーソル問題が起こらないこと、それ以外の機能は不要ということが分かった。調べてみると、Google Chrome は音声認識 API を標準で装備しており、Javascript プログラムで無料で利用できるとのことだった。IPtalk_s2t_Chrome ウィンドウもそれを利用していたのだが、こちらは認識結果を IPtalk に送っている。そうではなく、同じページの別枠に送るのなら、自分でも作れそうだしカーソル問題も起こらないと考え、アプリを作成した。

自作アプリの初期画面を図 1 に示す。下段は音声認識中の文字がリアルタイムで表示され、音声認識が終了したら上段に送られ修正ができる。他の機能はないので操作は簡単である。Windows PC で利用できる。



<https://www.deaf-s.tsukuba.ac.jp/sigika/webspeech.html>

図1 情報保障 Web アプリの初期画面と URL

3 対応ブラウザの増加

その後、Microsoft の Edge でも音声認識 API が装備され、本アプリが利用できるようになったが、Chrome と Edge では音声認識エンジンが異なるので、音声認識結果には違いが出てくる。どちらも語彙は豊富で認識精度は良いが、特徴としては、言い淀みや複数人の声を Chrome はカットする傾向があるが Edge は文字化する傾向がある、音声認識が確定するまでの文章が Chrome のほうが長くなることが多い、Edge では「。」や「?」が入るが Chrome は入らない、などが挙げられる。本アプリを試す際には、1 台の PC で Chrome と Edge の両方を開き、同時に音声認識をさせて、結果を比較して、どちらで利用するか判断してほしい。

4 置換語句登録機能の追加

初めのうちは認識精度の良さと使いやすさに満足し、少しの修正は苦にならなかったが、同じ誤変換を何度も修正させられた時にさすがに辟易し、修正パターンを置換語句として登録すれば自動修正される機能を追加した。これは、音声認識エンジンの語彙として辞書登録するわけではなく、特定の文字列を別の文字列に自動で置換するもので、IPtalk の「自訂」機能と同様の機能である。

5 講義情報保障における本アプリの効果検証

令和 6 年度 歯科技工科 3 年の矯正歯科技工学の

講義（3 校時×全 15 回、3 人の外部講師が交代で講義を担当、情報保障は筆者が担当）の第 10～12 回において、次の①～③の文字数を数えることにより本アプリの効果を検証した。

- ① 音声認識により入力された文字(※1)
 - ② 置換語句登録により修正した文字(※2)
 - ③ タイピングにより修正した文字
- ※1: 音声認識用マイク AmiVoice WT01 を講師が首掛けで装着し音声入力、Microsoft Edge 音声認識 API により文字化
- ※2: 第 1 回講義からの誤変換パターンを累積登録、「。」「?」は半角空白に置換するよう登録
その他：句読点や記号は検証対象から除外する。

検証方法を説明するための実例を表 1 に示す。音声認識と修正後の文字単位の差分をゴシック体で、置換登録語句を下線付きで示している。

表 1 本アプリでの講義情報保障の実例

音声認識	ギガスプリングリテイナーこれは <u>強制陽線</u> の <u>男性</u> を利用したホテル措置。		
	科学 <u>接種化学</u> の前田の保定装置です。		
	で <u>化学摂氏</u> の軽度の <u>創生</u> や <u>ねん</u> 店などの治療にも利用できる。これもセットアップすることが可能。		
修正後	次がスプリングリテイナー これは <u>矯正用線</u> の <u>弾性</u> を利用した保定装置		
	<u>下顎切歯</u> <u>下顎</u> の前だけの保定装置です		
	で <u>下顎切歯</u> の軽度の <u>叢生</u> や <u>捻転</u> などの治療にも利用できる これもセットアップでつくる事が可能		
		文字数	割合
① 音声認識による入力		65 字	71%
② 置換登録による修正		13 字	14%
③ タイピングによる修正		14 字	15%

表 1 から分かるように、誤変換は同音異義語が多い。特に、「下顎（かがく）」を「科学」または「化学」、「叢生（そうせい）」を「創生」など、専門用語

が一般的な語句に誤変換されることが多かった。

全3回の講義情報保障における検証結果を表2に示す。専門用語が多い内容にもかかわらず、音声認識で90%以上の文字が正しく表示できており、かなりの労力削減効果を得られたといえる。③のタイピングによる修正も、IMEに専門用語を登録したPCを使用していたので、同音異義語の修正はほとんどが変換キーだけで済んでおり、修正の労力は一からタイピングするよりは少なかった。

表2 3回の講義における本アプリの検証結果

	文字数	割合
① 音声認識による入力	55,240 字	94%
② 置換登録による修正	1,130 字	2%
③ タイピングによる修正	2,531 字	4%

6 音声認識のセキュリティ

サーバー型の音声認識は、提供元が音声データを保存し、音声認識技術向上のために再利用する場合があるので、個人情報を含む内容で利用する場合は注意が必要である。本アプリの場合、音声データの取り扱いはブラウザに依存する。

(1) Microsoft Edge の場合

「Web Speech API の利用において、音声データはいかなる目的でも保存されない」とプライバシーポリシーに明記されている。

(2) Google Chrome の場合

Web Speech API の音声データの扱いについて明確な記載は見つけられなかったが、Google プライバシーポリシーには、「ユーザーが承認した場合はサービス改善のために匿名化された音声データを保存し分析する」とあり、ブラウザのアカウント設定、音声設定により、音声データが保存され再利用される可能性がある。

※ 2024 年 12 月時点のプライバシーポリシーを確認（プライバシーポリシーは定期的に見直され改訂される。）

7 おわりに

本アプリは、誤変換をリアルタイムに修正することを目的として制作したものである。PC 1 台で表示

とリアルタイム修正も兼ねる仕様なので、画面を複製するためのディスプレイの準備があると良い。複数の端末を連携させるような機能はなく、それは欠点と言えるが、だからこそ、接続の手間がなく、使用に際してのハードルは低い。図1に記載のURLにアクセスすれば、インストール不要、アカウント不要ですぐに利用できる。特に他の人に使ってもらう際には、この簡便さが大きな利点だと感じている。

最後に、音声認識による情報保障の問題点を表3に、音声認識に適した音声入力環境を表4に挙げる。

表3 音声認識による情報保障の問題点

- 情報の精度が話し手や話し方に左右される。
- 言い淀みや冗長語も含む話し言葉は、文字にすると読みづらく、分かりにくいことがある。
- 要約を望む場合はニーズに合わない。
- 誤変換に困惑させられる。
- サーバー型音声認識の場合はインターネット接続が必要となる。

表4 音声認識に適した音声入力環境

- 雑音の少ない環境で行う。
- 音声認識に適した指向性マイクを講師が適切に装着する。
- 音声認識に適した話し方をする。
速すぎず遅すぎず、落ち着いた口調で明瞭に、文末を言い切る話し方 ⇒ 聞こえる人にも分かりやすい話し方を心掛けてもらえば良い。

この数年の音声認識の進化には利用するたびに驚かされる。今後さらに認識精度は上がっていくことが期待される。

【付記】

筑波大学附属聴覚特別支援学校研究倫理審査委員会の承認を得ている。

本アプリは、筑波大学 特別支援教育 教材・指導法 データベース ID 477 に登録されている。

【参考文献】

Microsoft Edge Privacy Whitepaper
Google プライバシーポリシー