

音声入力機器の導入に向けて

法人事務局

岡崎 美樹 谷畑 茂実

川連 寿恵 中迫 絵麻

津久井やまゆり園

押田 誠一郎 宮崎 剛志

1.はじめに

昨今の各施設の状況をみると、利用者の生活記録は個別支援に大変重要なものであり、且つ記録作成の時間確保が必要とされているが、職員の欠員が慢性化していることから、記録作成の時間確保が厳しい状況にある。

そこで、本活動のテーマ(活動の目的)として、

- (1) 利用者支援の時間、個人記録の時間を確保するために、音声入力の導入について検討し、会議記録作成等の効率化を図りたい。
- (2) 法人として、今後の検討課題にも挙げられていることから、音声入力機器を購入・試行し、次年度、本格導入に向けて検討したい。

以上の2つの目的を掲げて活動を進めてきた。

2.活動内容

活動の目的から以下の4点を活動内容として掲げた。

- (1) 音声入力機器の情報収集
- (2) 音声入力機器を使用しての評価
- (3) 導入に向けての機器の選定
- (4) 導入に向けてのシステム案作成

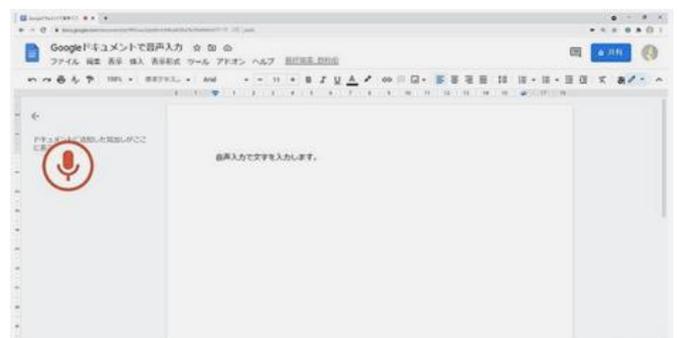
3.音声入力機器の情報収集

まず音声入力、いわゆる「文字起こし」と言っても、用途によっていろいろな種類のソフトがある。大きく分類すると、4つに分けることができる。

- ① 録音した音声ファイルを文字起こしするもの
 - ② リアルタイムで音声をテキストに変換するもの
 - ③ 画像や PDF の文章を読み取ってテキスト化するもの
 - ④ 手動の文字起こし作業をサポートするもの
- 記録の作成や会議録の作成を考えたとき、①②が法人として求めているものになる。こうした観点から音声入力機器について情報収集をした。昨今の音声入力機器の開発は日進月歩であり、開発が加速度的に進められている状況の中なので、情報は多岐にわたるため、収集した情報の中で主だったものを紹介する。

- (1) 「Googleドキュメント」の音声入力

下の画がパソコン上で開いた Googleドキュメントの画像になる。このドキュメントに音声入力のツールが組み込まれていて、赤いマイクのマークをマウスでクリックすると、音声を認識しテキストを自動で入力できる。



繰り返しになるが、Googleドキュメントはリアルタイムで音声を文字に変換してくれる Google が提供するドキュメントであり、マイクから入力された音

声をリアルタイムで文字に変換してくれる「音声入力」という機能が搭載されている。

このドキュメントの特徴だが、入力時間が長いと上手く音声を認識できないケースがある。長文を入力したい時は、ゆっくりはっきりと、息継ぎや句読点を打つタイミングで一旦音声入力をオフにするという方法がおすすめのような。更に、音声と音声の間に長い沈黙があると音声入力が「終了した」と判断して、一旦、音声入力が止まってしまうことが特徴である。

また、発言する人が頻繁に入れ替わると音声を上手く認識できないことがあり、音声入力が止まる恐れもある。

以上のことをまとめると、ウィークポイントとしては長い文章や沈黙、複数の声の聞き分けが苦手ということになる。会議の音声をそのまま議事録に起こすといった使い方は少し難しいのではないかと印象を受けた。

(2) UserLocal (音声議事録システム)

こちらは、株式会社ユーザーローカルが2020年9月15日から無償で提供しているサービスである。議事録作成を自動で作成できるだけでなく、テキストマイニング技術によって会話の流れを可視化することができる。

特徴として、複数の話者の音声を認識してテキスト化することができる。また、テキストマイニングにより、会議中にどんな話題が多かったのかを一目でわかるように表示してくれたり、さらにディープラーニング技術で感情やネガティブ、ポジティブを時系列で判定し、表示することができたりする。

下の図が実例になる。



「文字起こしエリア」に、時系列で会話をテキスト化した物が表示される。

「ワードクラウド」には、どのような単語が多く会話の中で使われているかを可視化して表示される。

「感情認識エリア」には、ネガティブ、ポジティブ等をグラフ化して表示される。

「単語出現頻度」もグラフとして分かりやすく表示される。

特徴としては、無料で使用できるという点にあり、メリットがある。しかし、Chrome ブラウザ上でのみでしか動作しない点はデメリットを感じた。

実際に無料でユーザー登録をして使用してみたが、誤変換がかなり多かった。リアルタイムで文字起こしをする点は魅力があり、テキストとして保存ができる点もメリットを感じた。

メリットとデメリットを総合すると、正確性に欠けることから実際に使用するには至らないと感じた。

(3) UDトーク

こちらは、コミュニケーション支援や、会話の見える化をするアプリである。とにかく共有を目的とした話し手のコミュニケーションをサポートするためのアプリであり、音声認識技術を使って、会話やスピーチをリアルタイムに文字化することができる。UD トークは、3つのコミュニケーションを実現している。



- ① 音声認識+音声合成
視覚聴覚障がい者間のコミュニケーションを叶える
- ② 多言語音声認識&翻訳
言語コミュニケーションを叶える
- ③ 漢字、かな、変換や手書き
世代間コミュニケーションを叶える

特徴としては、上記に上げた3つのコミュニケーションを実現していることである。

音声、キーボード、手書きと状況に合わせて入力方法を使い分けられることができることで、視覚聴覚障がい者とのコミュニケーションをスムーズにしている。また、議事録の作成や動画の字幕作成ができる。

150以上の多言語翻訳が可能で、双方向にリアルタイムでの多言語翻訳が可能となっている。

子どもとのコミュニケーションでは、初等教育から中等教育の学習レベルに応じた漢字の使用制限を設定することができる。

クラウドを利用したアプリケーションなので、スマートホンやタブレットだけではなく、PCをはじめとしたウェアラブルデバイスや、ARデバイスなど様々なデバイスで利用が可能となっている。

このアプリは、様々な企業や行政などの多くの場所で実際に使用されている。今後、期待できるアプリの一つだと感じた。

4.音声入力機器を使用しているの評価

(1) Google ドキュメントを使用し、別添ハンドアウト資料に沿った口述の様子を報告会で実演した。以下がハンドアウト資料になる。

はじめに

このハンドアウト(手元資料)は、ほとんどすべて iPad を用いて音声入力により作成しました。主に Google ドキュメントを使い、一通り入力を終えてから、文章の編集はカット&ペーストで行っています。

(1)端末について

なぜ iPad なのかと言うと、句読点や改行(←この部分は手入力です)、(かっこ)、「かぎかっこ」などをほぼストレスなく入力できるからです。

その他の端末を使った場合、例えば以下は Android スマホで入力したものです。

この文章は、Android スマホで入力してみました。

↑お分かりになりますでしょうか?しゃべった言葉が詰まると、次に入力を開始した時に半角の空白が入力されてしまいます。句読点は入力できますが、この微妙に空いた空白がその後の編集作業時にとっても邪魔になります。

そして、Chrome がインストールされた WindowsPC 上では Google ドキュメントのサイトを開くと、マイクからの音声入力が可能になります。

この文章は答弁 Windows PC で入力しました。メニューバーにあるツールボタンから 同点 音声入力を選択できます。

やはり、不要な空白が入ります。また、→、(読点)を「答弁」や「同点」と認識しています。ただし、法人内で使用されているノート PC に Chrome を簡単にインストールできることから、テストしながら音声入力を試してみるということはできると思っています。

(2)どのような場面で音声入力を行うのか (支援計画_個人記録_相談)

さて、今回 iPad での利用は、昨年度の場合、生活介護にて契約前の聞き取りを行い、契約を結び支援計画を作成する場合であったり、個人記録を作成するときに音声入力で行ってみました。また、今年度は相談を受ける立場であることから、対面でお話を聞く時に使ってみました。実際に音声入力を行い、ご本人には音声入力した文章を実際に見てもらおうことが出来、その場で同意をもらう場合には、とても重宝しました。

相談業務を行っている時、相手がお話している言葉をマイクが拾って入力されることがあり、2人程度であればそれぞれが話したことを記録することが出来るように思われます。

(3)会議など複数人が参加する場合はどうなのか?

iPad を使用した音声入力は、かなり優秀な機能を持っていると思われていますが、期間中、会議場面で使用する機会がなく、現状では未知数といえます。例えば、IC レコーダーを用いて、その内容を文字起こしするソフトや、アプリも存在しますが、有償の機能となっている場合が多く、今回の私が担当した部分では利用するには至りませんでした。ただし、(1)で取り上げたように、会議に参加している方が、それぞれ端末を持ち、Google ドキュメントを共有した上で、発言するというのであれば、一定の音声入力が可能になると思われています。また、1 台の端末を利用して音声入力をする場合は、マイクの性能(単一指向型ではないものなど)に依存すると思われています。

(4)プライバシーについて

今回の研究では、ハードウェア(端末)は別にして、無償のアプリを利用し、検証しています。従って、GAF A 等(Google、Amazon、Facebook、Apple ←手入力でした)の無償サービスを利用しているということで、基本的に入力した情報は、GAF A 側のクラウド(データセンターにあるサーバーが容量やサービスによって結びついていくこと)上にあるということ、一応、認識しておく必要があると思います。ただ、Zoom(中国)、Line(韓国)、また、レインボーネットも、どこにサーバーがあるのかは、私は知りませんが、原理的には同じことであり、そこはしょうがないのかなと思っております。

(5)さいごに

実は、私はキーボードでの入力が苦手で、四苦八苦して入力しては文章を推敲するということを行っていました。しかし、音声入力であれば、聞いたこと、答えたこと、思ったこと、考えたことをとりあえず口述で文章化することができます。その後再度推敲することで、入力してから文章を見て、前後関係を考え、表現を考えて行うということのみで、キーボード入力を省くことができるというのは、とても省力化につながるものになりました。実際、この文章を作るのにかかった時間は、およそ1時間程度でした。iPadを利用して、消費バッテリーは5%でした。

(2) iPad を使用したレインボーネット上の個人記録への音声入力の方法

iPad には、ブラウザアプリ「safari」に音声入力の機能が備わっているため、ブラウザ上で稼働するレインボーネットを「safari」で開くことでその機能が使用でき、直接入力画面に話した言葉が入力される。

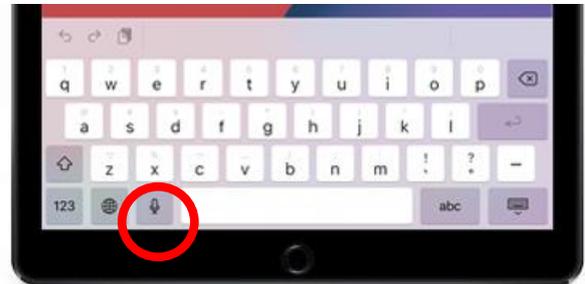
手順

- ①iPad を起動し、レインボーネットのアイコンをタップする。
- ②「safari」でレインボーネットに接続。ログインし、個人情報記録する画面に遷移する。

- ③マイクのボタンを押し、できるだけゆっくり・はっきりと話す。

図:iPad のキーボード

マイクボタンは下記の丸囲み部分



例)「今日は てん 活動に参加しました まる 改行 お菓子を食べて てん ユニットに戻っています まる」
と話した場合、実際は以下のように入力される。

**今日は、活動に参加しました。
お菓子を食べて、ユニットに戻っています。**

入力の精度については、入力された音声の質(ゆっくり・はっきり話しているか・周囲が騒がしい等雑音が入っていないか)により、左右されることはあるが、日常生活で使用する言葉についての誤変換は少ないように感じる。

5.今後に向けて

報告会のなかで実演をしたように、現在のパソコン環境下でも工夫次第で音声入力による文章作成は、ある程度可能である。しかし、使用できるデバイスの配備はもちろんのこと、デバイスのバージョンやマイクの性能などの環境に、入力の精度が左右されること、機器の使用についての手順の普及にも時間がかかると思われ、まだまだ手軽に誰でもが利用できるといった状況ではないと感じられた。

今後考えられる展開としては、

- 支援の合間に音声入力レインボーネット記録が作成できるデバイスとして iPad・iPad mini 等 iOS デバイスの配備強化

令和3年度 研究活動援助事業

- 会議録の音声入力の推進(Google ドキュメント・UDトーク・office365等の活用)
- レインボーネットの機能改修

が考えられるが、現在 ICT の技術がめまぐるしく進歩しているため、導入には費用や準備期間を要する。

今回報告した手法の導入には、まだまだデメリットが多いと思われるが、今後情報をよく注視し、導入までの期間等を含めて法人としての検討と推進が必要になってくるとと思われる。