

# 音声認識を情報保障支援に取り入れる前に

## 1. はじめに

AI技術の進展により、音声認識技術(以下、音声認識)を用いた情報保障支援の導入を進める大学も増えてきました。しかし、どの情報保障手段にも長所・短所があるように、音声認識を用いた支援にも、効果的な側面とそうでない側面があり、導入を検討する際には十分な準備と理解が必要になります。

本稿では、聴覚障害学生に対するリアルタイムの情報保障支援に音声認識を活用する際に、事前に知っていただきたいポイントについて解説します。情報保障手段にはさまざまなものがありますが、それぞれの利点や注意点を考慮しながら、選択肢の1つとして活用いただければと思います。

## 2. 音声認識を用いた情報保障支援の特徴

はじめに、音声認識を用いた情報保障について、メリットと留意点から、その特徴を概観していきたいと思います。

まず、聴覚障害学生への情報保障支援に音声認識を利用するメリットとして、以下のような点があげられます。特に、条件が良ければ、話された内容が自動で文字化されるという点は、最大のメリットで、多くの大学が音声認識を活用したいと考える理由もここにあるでしょう。

一方、情報保障として音声認識を利用する際には、留意しておくべき点も存在します。特に、認識率や表示される字幕の特性についての理解は、音声認識を利用していく上で欠かせないものなので、必ずご確認くださいということです。

また、音声認識を導入すれば、パソコンノートテイクなどの支援手段は不要になると思われている大学も多いようですが、実際には生成される字幕の質や活用可能な場面が異なるため、パソコンノートテイクの方が良い場面や音声認識では対応できない場面が存在します。このため、音声認識を利用した情報保障支援を始める際には、あくまで選択肢を1つ増やすつもりで取り組む必要があるという点にもご留意ください。

### メリット

- 一から手で入力する必要がなく、認識率がよければ、話し言葉をほぼそのまま字幕化できる。
- 認識率がよければ、修正にあまり手がかからず、支援者の負荷も少ない。
- 共通のアプリケーションを用いることで、認識された字幕を簡単に共有できるため、字幕を必要としている学生が複数在籍している場合でも活用しやすい。

## 留意点

- 話者や音声・通信環境によって認識率に大きな差が出る。
- 聴者の目から見て概ね正しく変換されていると感じる場合でも、聴覚障害学生にはそのことが判断できないので、正しく変換されているかどうかを確認するモニタリングや修正を行う支援者の配置が不可欠。
- 誤認識の中には、一見正しく変換されているように見えても、実は意味が通じていないものも含まれるので、支援者にこうした意識を持たせるための養成が不可欠。
- 話し言葉のままの字幕は、声の高さや強さ、抑揚などのパラ言語情報が失われている分、字幕だけで理解するのが難しいことがある。
- たとえ正しく変換されている文字であっても、字幕の量が多すぎると、読み切れなくなることもあり、学生によっては疲労度が大きい場合がある。
- 認識率があまりよくない場合や聴覚障害学生にとって負荷が大きすぎる場合には、音声認識を活用すること自体の見直しが必要になる。
- 音声認識があれば、パソコンノートテイクなどの支援は不要になると思われがちだが、実際には音声認識だけでは対応できなかつたり、パソコンノートテイク等を用いた方が良かったりする場面も多い点で留意が必要。

## わかりづらい誤認識の例

表示された字幕	本来の意味	説明
私は賛成ですけど。	私は賛成ですけど?!	文字だけでは怒っている雰囲気まで伝わらず、行き違いが起きてしまう
和歌山のどこ。	和歌山はどこ？	疑問文であるということが伝わらず、タイミングよく回答できない

以上のことから、情報保障の手段として音声認識を用いる際には、この方法が最適と言えるのかどうかの「見極め」や、モニタリングと修正のための「支援者」配置と養成、音声認識によって出力される字幕の特性(特に「読みづらさ」)に関する理解が不可欠と言えます。

- 支援手段としての「見極め」
- 支援者の配置と養成
- 字幕の「読みづらさ」への理解

次章から、これら3点の内容について、具体的に説明していきます。

### 3. 支援手段としての「見極め」

音声認識を情報保障支援に取り入れる際にもっとも重要になるのは、支援手段として適切かどうかの「見極め」です。他のマニュアル等でも触れている通り、音声認識の活用にはトラブルがつきもので、授業に導入していく前には十分なノウハウの習得が必要です。また、後述の通り、音声認識によって生成される字幕は、必ずしも読みやすいものではないので、本当に聴覚障害学生にとって適切な情報保障手段になるのかどうかは、慎重な判断が必要です。

このため、実際に利用を考える際には、少なくとも、以下のようなSTEPを踏んで、よくあるトラブルへの対処方法を学ぶとともに、音声認識が適している場面とそうでない場面の違いを理解してから、判断することをお勧めします。

#### 【STEP1】1対1の会話場面

はじめに利用をお勧めしたいのは、1対1の会話場面です。音声認識には、誤認識がつきものです。けれども、1対1の会話であれば、互いに内容を確認しながら会話を進められるので、意味が伝わりづらいところは、言い直したり、修正したりしながら話を進めることが可能です。

このため、まずは、支援室職員と聴覚障害学生の面談や、友人同士の会話など、互いに内容を確認しながらやり取りができる場面を選び、利用してみることをお勧めします。聴覚障害学生に勧める前に、職員同士で利用してみるのもいいですね。



#### 【STEP2】少人数での会話場面

1対1での会話に慣れたら、気心が知れた3~5人の会話で利用してみるといいでしょう。例えば、支援学生との話し合いや仲間同士の会話、あるいは支援室職員同士の打ち合わせ等です。

人数が増えると、音声認識を利用する上でのトラブルも増えます。誰かが認識開始のボタンを押すのを忘れて、ネットワークトラブルでなかなか文字が出てこなかったりすることもあるでしょう。でも、こうした「いい失敗」を繰り返すことで、音声認識を利用する際の留意点を学ぶことができます。



### 【STEP3】1方向の講義・講演場面

※情報保障としては  
利用しない試行運用

STEP 1~2を通して、音声認識の特性を掴んできたら、授業に近い場面で試行的に音声認識を稼働させてみるといいでしょう。でも、いきなり情報保障として利用するのは危険です。あくまでノウハウ蓄積の一環として、情報保障を必要とする人がいない場や別の情報保障手段が確保されている場での実践とすることが重要です。

例えば、FD 研修会等の場で試行的に運用してみたり、パソコンノートテイク等が配置されている授業の裏側で音声認識を立ち上げ、情報の量や質の違いを比較したりするとよいでしょう。

ただし、フォーマルな講演会の場合は、話者の話もクリアになりやすく、通常の授業よりも認識率が上がりがちな点は考慮に入れる必要があります。



以上のような STEP を通して、さまざまなノウハウが積みあがってくると、実際に利用を想定している場面で、本当に音声認識が活用できそうかどうか見えてくるのではないかと思います。そして、もし状況が許せば、ぜひ一度、導入を予定している授業の場で、音声認識を稼働させるとともに、その映像を 90 分間、音がない状態で見てみましょう。「これならしっかり勉強ができる！」と思える字幕になっているかどうか、それが「見極め」のポイントです。

### 【知ってください】

PEPNet-Japan では、大学の授業一特に、まだ情報保障を使い慣れていない学部 1～2 年生の授業に、音声認識による情報保障を導入することは推奨していません。それは、学生たちの多くが、高校までに文字通訳を通して「100%わかった」と思える体験を積み重ねておらず、本当に今の字幕で内容をつかめているのか判断する材料を持ち合わせていないからです。

また、誤認識やわかりづらい文章が出てきたときに、これは誤った情報だと判断するためには、ある程度の背景知識が必要です。そうした意味でも、学問の入り口に立ったばかりの1～2年生が音声認識を用いた情報保障を使いこなしていくためには、一定の時間が必要なのではないかと思います。

一方、ディスカッションなど、複数人が話す場面で音声認識を利用するためには、周囲の人々に働き掛ける力が求められます。表示された文の意味が分からない時、何らかのトラブルでアプリが止まってしまったときなどに、「ちょっと待って」「これどういう意味？」と言える力が必要だからです。

大学に入ってはじめて情報保障を使いはじめた学生達が、徐々に成長し、自分の力で周りに関わっていきたいと感じるようになった時、それが音声認識を導入する一つのタイミングなのかもしれません。これまでの実践の中では、こうした音声認識の特徴を生かして、上級生や大学院生のエンパワメントの機会に繋げている事例も報告されています。とても学びの多い取り組みなので、参考までにご覧いただければ幸いです。

#### 参考文献

第12回日本聴覚障害学生高等教育支援シンポジウム報告書

(セミナー2「音声認識技術を活用した情報保障—合理的配慮とエンパワメントの視点から—」) <http://hdl.handle.net/10460/1574>

## 4. モニタリングと修正のための支援者配置と養成

音声認識によって出力された字幕には、多かれ少なかれ何らかの誤認識が含まれています。また、たとえ誤認識がなかったとしても、話者の音声が聞こえない聴覚障害学生には、正しく字幕が表示されているかどうかを判断することができません。このため、音声認識を情報保障として用いる際には、出力された字幕をモニタリングし、間違いがあれば修正するとともに、正しく表示されていることを保障する支援者の存在が不可欠になります。

この支援者には、以下のような役割があります。

- 字幕をモニタリングし、正しく表示されているかを確認する
- 誤認識が含まれる場合には、その箇所を修正して再表示する
- 文字だけでは意図が伝わらない箇所があった時に、ことばを補ったり、修正したりする

このうち、3つめの意識を持つことは非常に重要です。

音声認識による字幕では、一見正しく変換されているように見えても、文字だけでは話者の意図が伝わらないことがあるからです。

たとえば、以下のような場面では、文字と手話どちらが重要と言っているのかを区別することができません。このため、支援者は意図が正しく伝わるように、ことばを補ったり、逆に削ったりする必要性が出てきます。

### 【出力された字幕】

聴覚障害者にとって、文字、手話によるコミュニケーションは非常に重要です

### 【考えるパターン】

- ① 文字や手話によるコミュニケーションは非常に重要(並列)
- ② 文字…(言い直して)手話によるコミュニケーションは非常に重要(訂正)
- ③ 文字だけでなく、手話によるコミュニケーションも非常に重要(付加・累加)
- ④ 文字よりも、手話によるコミュニケーションが非常に重要(比較強調)

しかし、音を聞きながら文字を見ていると、こうした違和感に気づきづらいものです。このため、モニタリングにあたる支援者には、しっかりと役割を伝え、適切に対応できるようトレーニングをしておく必要があります。

一般的に、パソコンノートテイクと比較して、音声認識による支援は、支援者の養成が容易だと思われるがちですが、押さえておくべきポイントを押さえられていないと、効果的な支援にならない場合もあるので、十分な注意が必要です。

## 5. 字幕の「読みづらさ」への理解

これまでも述べてきたように、音声認識を用いた字幕には「読みづらさ」がともないます。これは、①話し言葉をそのまま文字化しても読みやすい字幕にならないこと、②生成された字幕には誤認識が含まれていること、さらに③一度表示された文字が修正・再表示されることで、読みづらさを引き起こすことがあることなどが理由です。以下、それぞれについて順に見ていきましょう。

### 〇話し言葉を文字化することで生じる問題

音声認識ソフトは、条件が整えば、話し言葉をほぼそのまま字幕化することができます。これは一見素晴らしい利点のように思えますが、実は、そうとも言い切れないのが難しい点です。

ご承知のとおり、話し言葉には、声の高さや強さ、抑揚といった「パラ言語情報」が含まれます。聞こえる人々は、無意識のうちにこうした情報を取り入れながら、内容を理解していますが、文字にするとうちの情報が失われしまうので、意味がつかめなくなってしまうことも少なくありません。このため、話し言葉をいくら完全に文字起こししても、それだけでは意味の通る字幕にならないことがあるのです。

例えば、以下のような字幕で、スムーズに意味をつかむことはできるでしょうか？

障害学生への支援っていうのは、この間、第3次まとめが、みんな知ってますよね。文頭に、感動しましたが、価値だ。価値をもたらすって、こうほら、しっかり、大学教育にね。書かれています。

文章の途中で挿入句が入っていたり、呼応関係が消失していたりして、とても読みづらいのではないのでしょうか。そして、理解しようとする、自然と頭の中に音声が浮かんできませんか？これは、抜け落ちてしまった韻律情報を補完しようとしているためだと思われます。

けれども、普段、韻律情報をあまり活用していない聴覚障害学生にとっては、こうした情報の補完が難しいことがあります。このため、字幕のみでスムーズに理解できるようにするためには、以下のように文の前後関係を入れ替えたり、言葉を補ったりしていく必要があります。

障害学生への支援っていうのは…。この間、第3次まとめが(発表されましたが)、みんな知ってますよね。(文章を読んで)感動しましたが、文頭に(障害学生への支援は)「価値」だ。「(大学教育に)価値をもたらす」って、(書かれています。)ここに、ほら、しっかりと。

## ○誤認識から生じる問題

一方、音声認識の場合、先に述べたような問題に加えて、誤認識にともなう問題が生じます。

例えば、次のような字幕を見たとき、誤認識がどこに含まれているのか、また、元の言葉が何だったのかを想像することはできますか？

これから奨学生支援に就いての話をしますね。  
 大学とかの5高等教育機関で、ダイバーシティ&インクルージョンを考えると、今非常に重要になってくる考え方っていうのが。  
 奨学生支援だと思う。  
 です。  
 障害学生の支援っていうのは、2016年の4月に相が良いを理由とする差別の解消の促進に関する法律。  
 ですよ。



この中では、「就いて(ついて)」「5 高等教育機関(高等教育機関)」など、誤認識であることがわかりやすく、かつ、元の言葉も想像しやすいものもあります。

けれども、「相が良い(障害)」などは、誤認識だろうということはわかって、一見しただけでは何の言葉だったのかわからない可能性が高いでしょう。

さらに、「奨学生(障害学生)」などは、もっともらしく変換されてしまっているので、誤認識であることすら気づけない場合があるかもしれません。

また、句読点が誤って挿入されたり、ちょっとした情報が抜け落ちたりしてしまうのも、音声認識でよく見られる特徴です。たとえば、次頁の図では、4～6行目が細切れに改行されています。これは元々「考え方っていうのが、障害学生支援だと思うんです」という一連の文章でした。この部分、音を聞きながら文字を見ていると、これらがただの句読点の誤りだということがわかります。しかし、文字だけを見ていると、ここに何か情報があったようにも見えて、不安に感じてしまいます。

一方、8～9行目も、同じように途中で文章が改行されただけのように見えますが、実は最後の文では「有名ですよ」と話されていました。些細な情報に過ぎないと思うかもしれませんが、講義の中で話されている法律について知らなかった学生にとっては、「覚えておかなきゃ」と思うきっかけになる重要な言葉の一つです。



- 1 これから奨学生支援に就いての話をしませぬ。
- 2 大学とかの5高等教育機関で、ダイバーシティ&インク
- 3 ルージョンを考えると、今非常に重要になってくる
- 4 考え方っていうのが。
- 5 奨学生支援だと思っ。
- 6 です。
- 7 障害学生の支援っていうのは、2016年の4月に相が良
- 8 いを理由とする差別の解消の促進に関する法律。
- 9 ですよね。

情報が抜けた??

途中で改行されただけかな？

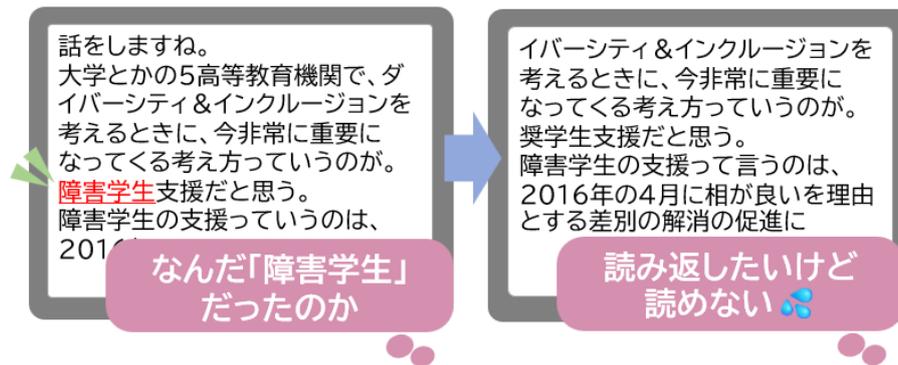
## ○修正から生じる問題

これまでに述べてきたような問題を解決するためには、読みづらい箇所や誤認識を修正する支援者の配置が重要であることがわかるでしょう。しかし、この修正もまた、音声認識による字幕の読みづらさを引き起こす一因になってしまうことも忘れてはいけません。

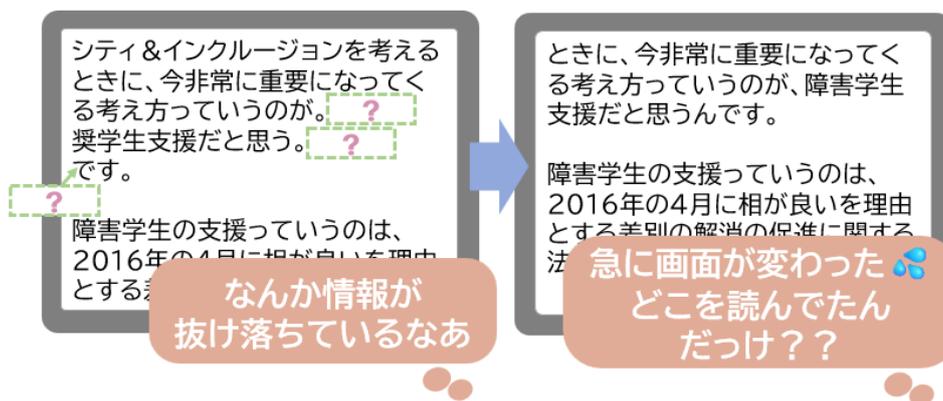
というのも、音声認識による字幕は、その特性上、一旦、認識された文字が表示された後、修正・再表示されるのが一般的です。アプリによっては、修正が済むまで文字を表示しない形になっていたり、表示のタイミングを切り替えたりできるものもありますが、こうすると字幕が表示されるまでにかなりの遅延が生じてしまいます。この遅延は 20~30 秒程度、場合によってはもっと長くなってしまいうこともあります。これでは、周りの状況についていけないので、多くの場合、先に修正前の文字を出し、あとから修正・再表示する方法がとられているのです。

しかし、この方法だと、一度読んだ文が修正される形になるため、修正が反映されるたびに、どこが修正されたのかを探して、内容を読み返す作業が発生してしまいます。たとえば、先に示した文章でも、「奨学生」が「障害学生」だったとわかった時点で、改めてはじめてから文章を読み返したくなるでしょう。

これが、誤認識が表示されたすぐ後に訂正されるのならまだいいのですが、多くの場合、ある程度字幕が進んでから前の部分が修正される形になります。そうすると、3~4行戻って読み返さなければいけなかったり、読み返そうと思った部分がスクロールされて画面の外に消えてしまったりして、ストレスに感じることも少なくありません。



また、現状の音声認識では、不自然な位置に改行が入ることも多いので、文の途中で改行されたり、逆に2つ以上の文が連なって1つの文として表示されたりしてしまいます。さらに、ソフトウェアによっては、改行を削除して文章を繋げると、画面上の文字配置が大きく変わってしまって、読んでいた箇所を見失ってしまうことがあります。



このように見えてくると、一見、多くの情報を文字化できているように見える字幕でも、実は聴覚障害学生が正しく受け取れる情報は、思っている以上に少ない可能性があることがわかります。聴覚障害学生の中には、それでも「だいたいわかるからいい」という場合があるかもしれませんが、わからない部分を想像で埋めながら授業に参加することが、本当に「教育の保障」と言えるのか？今一度、考えていく必要があるのではないかと思います。



音声認識を利用した情報保障に関する参考文献として、ぜひこちらもご参照ください。

- 第15回日本聴覚障害学生高等教育支援シンポジウム報告書  
前日特別企画「支援技術のさらに効果的な利用に向けて」  
<http://hdl.handle.net/10460/00002039>
- 第12回日本聴覚障害学生高等教育支援シンポジウム報告書  
セミナー2「音声認識技術を活用した情報保障—合理的配慮とエンパワメントの視点から—」  
<http://hdl.handle.net/10460/1574>
- 第16回日本聴覚障害学生高等教育支援シンポジウム  
企画1「聴覚障害学生を理解する—教育背景と心理から—」  
<http://hdl.handle.net/10460/00002101>

- 発行日：2021年8月5日
- 更新日：2024年5月29日(第2版)
- 執筆：白澤 麻弓(筑波技術大学 障害者高等教育研究支援センター)  
萩原 彩子(筑波技術大学 障害者高等教育研究支援センター)
- 編集：日本聴覚障害学生高等教育支援ネットワーク(PEPNet-Japan)事務局
- 発行：筑波技術大学 障害者高等教育研究支援センター  
〒305-8520 茨城県つくば市天久保 4-3-15 URL <https://www.pepnet-j.org>

※本事業は、筑波技術大学「聴覚障害学生支援・大学間  
コラボレーションスキーム構築事業」の活動の一部です。



国立大学法人  
筑波技術大学

PEPNet-Japan