

# 音声認識ソフトを用いた講義の支援方法に関する検討 第2報

## － 導入上の課題点と利用可能性 －

### Introducing Speech Recognition Software to Help Students with Hearing Difficulties

西村知恵 樋田康宏

Tomoe NISHIMURA Yasuhiro HIDA

札幌学院大学 電子計算機センター

Computer Center, Sapporo Gakuin University

Email:r120211@e.sgu.ac.jp

概要：札幌学院大学電子計算機センターのサポートデスクは、学生スタッフが教職員と協働して、学生や教員に対するPC利活用相談業務を行っている。この業務の一環として、聴覚障害学生向けに映像教材の字幕入れサービスを行い講義時の情報保障の支援を行っているが、モバイル機器の普及や音声認識ソフトの性能向上をふまえ、別の側面からの講義支援も可能であると考えた。今回、音声認識ソフトを用いて教員の発話を文字化し教室内におけるノートテイク支援の補助手段として用いることの課題点と、講義における利用可能性を検討し、その結果を報告する。

#### 1. はじめに

本学では聴覚障害学生の講義支援として「ノートテイク」と呼ばれる支援を行っている。これはボランティア学生(以下、テイカーと呼称)2人が聴覚障害学生と共に講義を受講し、ノートまたはPC上に教員の話した言葉をそのまま文字化するという支援である。聴覚障害学生が講義を受講するためにはなくてはならない支援だが、一方でテイカーには速記やタイピング速度などの資質が求められるだけでなく、90分の講義を文字化する作業は大変な負担にもなっている。こうしたテイカーの負担を軽減し、かつ聴覚障害学生が講義を受講しやすくするための工夫として、音声認識ソフトをノートテイク支援の補助手段として用いることを検討した。

#### 2. これまでの取り組みと課題

本研究におけるこれまでの検討<sup>(3)</sup>を踏まえると、音声認識ソフトをより高い精度で利用するためには使用機材の状態を講義に合わせるよりも、講義および発話する教員の状態を使用機材に合わせる方がよいと考えられる。今回はそのひとつとして、話者の発話速度による認識精度への影響について検討した。

#### 3. 発話速度による認識率の検証の手順

本実験には、パソコン上で動作する音声認識ソフト「ドラゴンスピーチ 11J」を用いた。雑音の入らない静かな室内環境でヘッドセットマイクを接続したノートパソコンに発話し、それを認識させた。手順としては、まず、ドラゴンスピーチにトレーニング(ソフトに使用者の声や話し方の癖を覚えさせ、認識精度を上げるための設定作業)を行った後、文献<sup>(1)</sup>より抜粋した3441文字(句読点など記号を抜いた場合3085文字)の文章を約10分間で読み上げた。これはアナウンサーの話す平均速度といわれており、人間が聞き取りやすいといわれる「1分間に300～350文字」という基準<sup>(2)</sup>に則ったものである。以上の条件で読み上げられたものを基準とし、同じ文を基準より速く読み上げたもの(約8分)と、遅く読み上げたもの(約11分半)、計3パターンを記録した。そして各条件ごとに所要時間、音声認識後の全体の文字数、誤入力数、読み方は同じだが変換が間違っている文字数の4つを分析し、それら4項目から正答率を算出し、比較した。また、講義の状態に近づけるため、「まる」「かいぎょう」など記号を発話せず、

「教科書読み」の状態でも音読した。さらにコンピューターに話すように丁寧に話すことを意識せず、普通に語りかける状態で話すよう気をつけた。

#### 4. 実験結果

上記の条件で実験を行った結果を示す。表1は3085字の文章を、1分間に発話する文字数を変えて同じ文を認識させた場合の結果である。表1にある全体の文字数は音声認識結果の文字数であり、元の文章に比べて字数が減っているのは、元の文章では平仮名で表記されている部分に変換されていたためである。誤入力は発話と完全に異なる文字列、変換が違う文字数は読みが同じだが意味が異なってしまいう変換が行われていたものをカウントした。正答率は「 $(① - (② + ③)) / 3085$ 」という計算式で算出した。その結果、認識精度は1分間当たりの発話字数が減ることに比例して上昇することが判明した。

表1 実験結果

	A(遅い)	B(基準)	C(早い)
所要時間	11分25秒	9分55秒	7分52秒
1分間当たりの発話字数	268文字	308文字	385文字
全体の字数①	3024	3014	3010
誤入力字数②	102	147	268
変換が違う字数③	31	24	27
正答率	89.4%	86.6%	78.4%

#### 5. 考察・課題

発話の速度に留意を行うだけで、90%近く正確な認識を行うことが可能だと判明した。これは発話速度を遅くすることにより、言葉をより丁寧に発話するようになったことの影響によるものと考えられる。

また、今回判明した問題点がいくつかある。まず1つは、内容が多く、進行速度が速い講義には向い

ていない点である。今回の実験結果からわかるように発話速度は認識精度に影響する。誤入力の回数によっては逆に文字情報が聴覚障害学生の理解を妨げる可能性もあるため、そうした講義への導入は望ましくないと考えられる。また教員に対して早口にならないよう協力を要請するなど、音声認識の特性の理解を促す必要がある。もう1つの問題点は、会話表現を用いて話すだけでは句読点や改行が行われないことである。文中で適宜「まる」「てん」「かいぎょう」などと発声しなければ句読点や改行が行われないままに文章が続いてしまう。そのままでは非常に読み辛く意味の切れ目もわからないため、文章だけを読んでいると意味の取り違えが起きる可能性もある。また誤変換によっても同様に意味の取り違えが起きる可能性があるため、適宜改行や句読点の入力、変換を修正するための補助人員についてもらうなどして改善する必要がある。こうした問題点をクリアすることができれば、実際の講義に用いることも可能であると考えられる。

#### 6. おわりに

聴覚障害学生への講義支援の補助手段として音声認識ソフトを用いるため、音声認識の特徴を捉えるための調査の一つとして、音声認識ソフトが発話者の発話速度にどの程度影響されるかについて検討した。1分間当たりの発話文字数を変えた際の誤入力数、全体の正答率などを示し、発話速度は認識精度に影響をもたらすことを確認した。また、こうした結果から講義へ用いる際の問題点を検討した。

##### 参考資料

- (1)香山リカ(2004)『生きづらい<私>たち 心に穴が空いている』講談社。より、40~46ページを抜粋。
- (2)NHK アナウンスルーム Q&A  
<http://www9.nhk.or.jp/a-room/qa/>  
 札幌学院大学サポートデスクホームページ  
<http://ext-web.edu.sgu.ac.jp/supportdesk/top.html>
- (3)船山夏那、樋田康宏:”音声認識ソフトを用いた講義の支援方法に関する検討”,PCカンファレンス北海道2013論文集,(2013)