

音声認識ソフトを用いたノートテイク代替支援の可能性に関する1考察

皆川雅章^{*1}

Email:minagawa@sgu.ac.jp

*1: 札幌学院大学社会情報学部社会情報学科

1. はじめに

演習型講義において、音声認識ソフトを使って教員の口頭説明を文字化して表示し、教室内の聴覚障がい学生に情報保障を試みた結果を報告する。

本学では、聴覚障がいを持つ学生(以下、対象学生)に対して、講義中に本学学生による「ノートテイク」あるいは「パソコン通訳」の支援⁽¹⁾が行われている。これによって講義中、対象学生の横に支援を行う学生(以下、支援学生)が着席し、教員が話す言葉をノートに筆記、あるいはタイピングによりPC画面上に表示し、対象学生はそれを見ながら講義を受けることができる。

昨年度、この支援と並行して、PCを用いた演習型講義での説明や指示などを、音声認識ソフトを用いて文字化し、対象学生に伝えた。これらの講義においても支援学生が配置されているが、実質的にテイクを必要とするのは、講義の始めに行う、演習に関する説明や指示、演習中の補足説明の部分に限定されているので、教員だけに対応できないかと考えて作業代替方法を探った。タブレット型端末やスマートフォン上の音声認識ソフトを使用して、この支援を2つの講義で実施し、一定の成果を得たが、音声の変換結果、変換スピード、対象学生への文字情報提供方法など、実際の教室で円滑な活用を行うためには、検討・解決すべき課題が見つかった。

本報告では、講義中に行った説明を実験用例文として用い、デバイスやアプリケーションソフトが異なる環境における、音声認識ソフトの性能、デバイスの操作性、誤変換の修正法、リアルタイム性等を比較実験した結果を示し、ノートテイクによる情報保障の代替支援として音声認識ソフトを導入する際の課題を考察する。以下では音声認識の結果が文字情報に置き換えられることを「変換」、音声認識ソフトの処理結果を「変換結果」と呼ぶことにする。

2. 試行結果概要

2012年後期の2つの実習科目における実践例を示す。1つはHTMLの基礎を学ぶ「情報デザイン基礎」、もう1つは3DCGの制作方法を学ぶ「CG制作論」である。いずれもPC実習室での説明と演習が並行して実施される講義形式を採用している。

これらの科目では毎講義時にテキストをWeb上に掲載し、その内容に沿って履修者は例題と演習課題に取り組む。また、対象学生は「ノートテイク」による情報保障を受け、教員の説明は対象学生の隣に座っているノートテイクの学生によって文字化される。さらに対象学生に対しては、別途Web上のテキストを印刷して配布している。

当初の試みではタブレット型端末iPad上で音声認識ソフト「Dragon Dictation」⁽²⁾を用いた。端末に向かって発話すると認識結果が図1のように表示され、これをスクリーンに投影する。

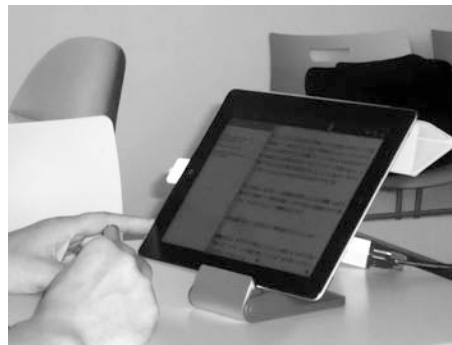


図1 タブレット型端末と音声認識結果



図2 講義時の説明の様子(CG制作論)



図3 講義時の説明の様子(情報デザイン基礎)

図2ではCG制作論においてWeb上のテキストを表示させたスクリーンの右横に、変換結果を表示している。図3は情報デザイン基礎において、教員が立った状態で端末を操作しながら説明し、対象学生(前方左側)がその説明を聴いている様子を示している。この端末の場合、片手で持って操作することが可能なので、誤変

換をその場でソフトウェアキーボードを用いて訂正できる。

次の試みとして、スマートフォンの音声認識機能を用いた。ここでは、スマートフォン接続したPC用モニタを対象学生の近くに配置した。授業後に変換結果のテキストデータを学生にメールで送付した。

3. アンケートによる評価

上記の2つの演習科目のテイク各1名にアンケートを実施した。アンケートの内容を以下に示す。

アンケートの内容 (関連部分のみ)

Q3) ノートテイクが (都合がつかない等の理由で) 不在のときに、音声認識ソフトがどの程度テイクの代わりを出来ると思いますか？
 Q4) 音声認識ソフトを使った説明は、通常の講義ではどの程度活用できると思いますか？活用できそうな場面、クリアすべき問題点は何かと思いますか？
 Q5) 音声認識ソフトを使用した文字表示に対する評価・感想を書いてください。(どのような内容、観点からでもけっこうです。)

回答結果は次の通りである。回答AはCG制作論、Bは情報デザイン基礎論のテイクによるものである。

Q3) の回答

回答A) 音声認識ソフトがいつもどおりにきちんと働けば代わりはできると思います。しかし、まれにおこる動かなくなることや上手に声をひろってくれないことがあったときは難しいと思います。
 回答B) ノートに書いてある内容と同じことが表示されていたので、だいたいカバーできると思います。

Q4) の回答

回答A) 十分に活用できると思います。声を認識してもらうためにゆっくり話さなければならぬことが多いので、その間、他の学生が困ってしまうことがあるかもしれません。
 回答B) 今回のような、各自で演習をする講義などで説明をする場面では、とても活用できると思います。表示する画面がスライドの近くにあると同じ視点で平行して見れるので、わかりやすいです。(テイクでは、テイクとスライドを交互に見る必要がある。) 言葉の変換スピード、変換ミスが減るといいと思います。

Q5) の回答

回答A) ほとんど正確に文字を読み取ってくれていてすごく良いと思います。この最後の講義で使用しているパソコン型のものに表示されるのが一番被テイクさんにとって見やすいのではないかと感じました。(スクリーンは上を見なければならぬ、メールは毎回ひらかなければならぬ所が大変かなと感じました。)
 回答B) 変換ミスがないと、ノートよりも速く表示されていた

ので、途中からはほとんどテイクが必要ないかなとも思いました。すごいと驚きました。変換ミスがないことと、指示語(「これ」、「あれ」、「ここ」などと具体的に説明するようにしてもらえば(「スライドの青い部分」とか)、ほぼテイクがなくてもカバーできると思います。

Q3 については、機器のネットワーク接続の状態と話者の音声の状態に起因すると思われるが、講義によっては変換結果が安定しないことが発生したために、機器の動作の安定性を懸念する意見が出されている。変換結果が安定している場合には、ノートテイクの筆記とほぼ同じ内容が得られている。

Q4 は、90 分間話し続けるような講義の場合に、今回の方法がどの程度機能するかを尋ねたものである。回答 A では、対象学生向けに講義を行った場合、他の学生にとっては、講義の進行上、不満が生じることを指摘している。回答 B においても、本方法の適用可能性が講義の形式に依存することが述べられている。

Q5 は対象学生が文字情報を見やすいようにする工夫の必要性、説明上の工夫の必要性が指摘されている。

4. 課題点

これまでの取組みを通じて、以下の6点を本アプローチの課題として挙げた³⁾。

課題1：リアルタイム性の問題

講義の流れを壊さないようにするために、発話から文字化までの所要時間が短いことが要求される。今回は発話後に音声認識を行い変換結果が文字化されるまでに5秒~10秒程度の時間がかかり、その都度待ち時間が発生することで、講義の流れに影響を与えた。説明する側も変換結果が表示されるまで待つ必要がある。

課題2：誤認識を起こさないような発話の工夫

正確な音声認識結果を得るには、発話の工夫を必要とした。発音が曖昧な場合、特殊な用語を使う場合には、全く関連のない言葉に変換されることが起こる。一語ずつ意識して明確に発音することはもとより、認識されやすい言葉への置き換え、同音異義語がある場合には言葉の前後関係に基づいて音声認識させるために文脈を意識した発話なども行う必要がある。

課題3：誤変換結果の速やかな訂正

タブレット型端末/スマートフォンではソフトウェアキーボードを使い、片手で持った状態で文字の編集が出来るので、誤変換された場合にはその場で訂正できるが、変換時の待ち時間に加えて訂正時間が発生するので、説明の流れを壊さないためには、訂正作業を速やかに行う訓練が必要となる。

課題4：対象学生への文字情報の提示方法

スクリーンを使用する場合、演習中などは説明が行われていることに気付かない場合が生じた。モニターを対象学生の近くに置くなど、視線を大きく動かさずに文字情報を見ることが出来るようにする等の工夫が必要である。

課題5：講義全体の運営の問題

テイクの配置を前提とした講義では、発話のスピードや発音など、筆記のしやすさを意識した説明を行

うとしても、他の受講生は対象学生とテイカーの存在を強く意識することは少ない。今回の方法で待ち時間が生じると、他の受講者は、講義の進行が遅いと感じることは明らかで、教室内の緊張感の維持が難しくなる。講義時間の大半が教員の口頭説明となるような形式の講義では、その問題が顕著になると考えられる。

課題6：機器とソフトウェアの性能・利用環境

昨年度の前半は、タブレット型端末と Dragon Dictation、後半はスマートフォンと内蔵音声認識機能を用いた。使い勝手という観点から①機器のサイズと重量、②ソフトウェアキーボードのサイズ、動作上の問題から③通信環境、は確認する必要がある。①、②については利用者によって意見が異なるが、③については重要で、通信が不安定になる、あるいは切断されることによって、講義の途中で機器の使用を前提とした説明が成り立たなくなる場合もある。

5. 実験

今回は、使用する機器とソフトウェアによる音声認識性能、変換結果の訂正の容易さを比較検討することを通じ上記課題へのアプローチを図った。

5. 1 使用機器・ソフトウェア

上述の機器・ソフトウェアに加え PC 上で動作する音声認識ソフトを使用した。PC を使用する場合、ネットワーク接続は不要である。

装置1：スマートフォン+音声認識ソフト(DD)

装置2：タブレット型端末+音声認識ソフト(DD)

装置3：PC+音声認識ソフトA(AV)

(DD:Dragon Dictation, AV:Ami Voice)

スマートフォンとタブレット型端末の場合、音声入力後に、逐次的に認識を実行させるための画面タップの操作が入るが、PCの場合、自動的に継続して変換がおこなわれる。図4は今回使用したAmi Voice⁽⁴⁾の画面である。ヘッドセットからの音声入力が逐次認識されて文字に変換され、表示されていく。

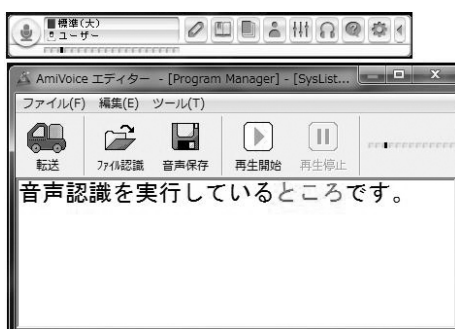


図4 Ami Voice の画面

5. 2 実験用例文

次の例文を用い、上記3つの装置を使った音声認識結果の比較を行う。この文章は前述の講義中の説明の1回分に部分的に変更・削除を行ったものである(文字数1256、以下では「例文」とする)。

実験用例文

おはようございます。

最初に色の指定方法について復習しておきましょう。RGBすなわち、赤、緑、青についてその明るさを指定しました。指定する方法は色の名前を指定する方法と、数値で指定する方法がありました。数値で指定する方法は10進数で指定する方法、それから16進数を使う方法があります。ここでは16進数を使い、色の指定は16進数2桁で表します。赤、緑、青のそれぞれについて指定します。16進数2桁を使いますので、3つの色を合わせると6桁になります。ここをみてください。この色は赤だということがわかりますね。色の指定はFF0000となっています。最初のFFが赤の明るさです。それから次の00これは緑の明るさです。そして、最後の00は青の明るさです。全体では赤の色だけが存在していますのでこの色になりますね。

2番目の部分を見てみます。今度は、真ん中の緑の部分だけがFFで、それ以外は00になっています。従ってこの部分には緑色だけが存在します。最後のこの場合は、青だけが存在していることとなります。そして、いずれの色も00ではない場合には、赤と緑と青が混ざった色が作られます。1番目の例題で、16進数で色を指定する方法を確かめてみましょう。

この図を見て下さい。赤、緑、青の順に色の指定はどうなっていますか。先頭の赤の部分はすべてFFとなっています。そして緑と青の部分が変わっています。これは赤の色を固定して緑と青を段階的に変化させています。縦に緑の変化が書かれています。横は青の変化です。このようにして赤と緑と青の組み合わせでどのような色が作られるかが分かります。

こちらが色の指定です。1行目の先頭がFF、2つ目の緑の部分もFFです。そして3つ目の青の部分だけが大きな値から小さな値に変わっています。FFから始まってCC、99、66、33、00と段階的に小さくなっていきます。その変化がこの部分に相当します。

2つ目も見てみましょう。真ん中の緑の部分がCCで固定されています。そして先ほどと同様、緑がFFから順に小さな値となっていきます。以下同様にして、真ん中の緑の部分を段階的に下げて行きます。その繰り返しでこのような表ができます。

あとはこの表を見て赤と緑と青の組み合わせによって色が変化することを確認してください。

赤の部分をFF以外の値にすると、別な色の表ができますが、それについては、今回は省略します。ここでは、このように赤がFFの場合について表を作ってみましょう。これ以降は色の組み合わせの説明です。16進数を使って色を指定する方法は既に説明しましたね。ここでは三つの色を組み合わせます。

この図を横にみてください。横に並べた3つの組み合わせが配色をする上でエレガントに見える組み合わせです。この3つ以外にも、このような組み合わせが考えられます。

この下にあるのは楽しさを連想させる色の組み合わせです。例えば幼児向けのおもちゃなどのホームページでは、このような色の組み合わせが見られます。ここにそれ以外の色の組み合わせ例を載せてありますので、必要に応じて活用してください。

こちらが課題です。テキストにあるような配色を使ってみましょう。

課題で不明な点はありませんか。

5. 3 実験結果

上記の装置1~3を用いて例文の認識・文字情報への変換を3回行った結果を示す。表1の誤変換数は、「発話したものが例文通りに文字に変換されなかった」言葉の個数である。同一例文を同一装置を用いて認識している場合でも、発話の仕方によって結果が異なっている。表2の変換時間(訂正なし)は例文を発話し、文字に変換され終わるまでの所要時間である。通常の会話を想定した速さで読み、装置1と2の場合には、ほぼ一文ごとに変換操作(画面のタップ)が行われている。そのため、変換時間は装置1に比べて長くなっている。なお、今回、装置1・2を使うに際して、ネットワーク接続は常時高速に利用できる環境にあったので、試行結果のところでも述べたような、通信時間に起因する処理の遅れは発生しなかったと考えている。表3の変換時間(訂正あり)は、例文と変換結果との間に違いがある場合に、その都度各機器・ソフトウェアの機能を用いて訂正を行ない、例文通りに修正するまでに要した時間である。装置1・2ではソフトウェアキーボードを使っている。表4の特徴的な誤変換の例は、同一装置において複数回発生した誤変換である。

表1 誤変換数

	1回目	2回目	3回目	平均
装置1	33	30	22	28.3
装置2	23	39	30	30.7
装置3	22	23	19	21.3

表2 変換時間(訂正なし)

	1回目	2回目	3回目
装置1	6分58秒	6分39秒	6分32秒
装置2	8分49秒	6分29秒	6分21秒
装置3	6分19秒	5分50秒	5分45秒

表3 変換時間(訂正あり)

装置1	13分45秒
装置2	14分59秒
装置3	12分29秒

表4 特徴的な誤変換の例

装置1	先頭→戦闘(3), 3つの→光の(7), 例題→0代(3), 値→体(3), 青だけ→青竹(3), 縦→館(2), 表→氷(3), 表が→評価(2)
装置2	値→体(5), 3つの→光の(3), 表が→評価(4), 青だけ→青竹(2), 縦→館(2), 先頭→戦闘(2), 例題→0代(2), FF→てふてふ(2)
装置3	配色→退職(3), 会食(2), 対処, 2つ目→た爪(2)

(注: 括弧内は発生件数)

6 考察・課題点

誤変換数の平均値は装置3の場合の値が小さくなっている。表4の特徴的な誤変換の例から考えると、同じ変換のパターンの繰り返しが誤変換数の増加に影響している可能性がある。話者に依存するという見方も可能であるが、今回は複数話者による実験は行っていない。現段階では変換の特性を把握し、正しく認識されやすいように発話の方法を工夫して変えるなどの解

決策が考えられる。

変換時間については、装置1・2と装置3との間で差が生じているが、認識結果を文字情報に変換するためのタップ操作の影響がある。いずれの装置においても、変換が終わるまで待ち時間は感じられなかった。なお、今回は装置1・2のネットワーク接続状態は良好であった。このタップ操作を介することの影響は、教室内のやりとりに時間的な余裕がある場合には許容範囲と思われる。例えば、ゼミナールなどにおいて、意見交換を行うような場合⁽³⁾には、有効と思われる。

誤変換の訂正は、いずれの装置でも同様の操作(訂正箇所の確認、訂正箇所へのカーソル移動、キーによる修正入力)を行う必要があり、操作上の優位性に大きな差は見いだせなかった。ただし、装置1,2は片手に持って操作できるので、講義中の発話時に教員が教室を移動するような場合に有利で、教壇等の定位置にいる場合には装置3が向いていると考えている。

同一の例文に対して、誤変換数に約2倍の違いが生じているが、この差が装置の性能に起因するのか、使い方(発話の方法)に起因するのかを判断するのは現段階では難しく、ソフトウェアのチューニングによる性能向上の可能性を含め、さらなる実験に基づく検証を必要とする。発話の工夫で誤変換を減らす余地があると考えており、例文の文字数程度の文章で10個以内の誤変換数であれば、訂正時間が発生しても講義の流れへの影響は小さいと予想している。

音声認識ソフトを用いてノートテイクの代替支援を行うことについて、今回のような講義環境においては、上述のような工夫を行えば可能であると考えている。また、教室外での対象学生の個別の修学相談等において、担当教職員に手話のスキルがなくても発話を文字情報として伝達できるので、より正確な意思疎通のために活用できる可能性がある。

7 おわりに

Web ページ作成などの演習型講義において、音声認識ソフトを使って講義中の教員の口頭説明を文字化して表示し、教室内の聴覚障がい学生に情報保障を行う試みを報告した。講義中に行った説明を実験用例文として用い、デバイスやアプリケーションソフトが異なる環境における、音声認識ソフトの性能、デバイスの操作性、誤変換の修正法、リアルタイム性等を比較実験した結果を示し、ノートテイクによる情報保障の代替支援として音声認識ソフトを導入する際の課題を考察した。

参考文献・URL

- (1) 本学で実施されているノートテイクに関して
札幌学院大学バリアフリー委員会
<http://bfc.sgu.ac.jp/>
- (2) Dragon Dictation に関して
NUANCE <http://www.nuance.com/>
- (3) 皆川: "演習型講義における音声認識ソフト活用の試み", 教育システム情報学会研究報告研究報告, Vol. 27, No. 6 (2013)
- (4) Ami Voice に関して
アドバンスド・メディア <http://sp.advanced-media.co.jp/>