

情報提示と理解度の相関関係に関する実証研究 — 音声認識技術の有効活用のために —

岡田 菜穂子
森 岡 龍 治
須 藤 邦 彦

要旨

障害等のある学生への情報保障として音声認識技術の活用が進んでいる。音声認識技術を有効活用するためには誤変換の対応が重要となる。同時に授業等の情報は音声のみではないことから、他の資料などの情報を的確に組み合わせることで授業情報の理解度は向上するはずである。

本稿では、情報の提示の仕方と理解度の相関関係を確認するためにアンケート形式の検証を実施した。アンケート調査結果と音声認識結果とを組み合わせ分析を行ったところ、情報量と理解度は比例する傾向にはあるものの絶対ではなく、正確な情報を適切な方法で提示する必要性が明らかとなった。加えて、有効な情報提示は、誤認識を補正するためにも有効であると指摘できる。

キーワード

高等教育機関, 障害学生, 修学支援, 音声認識, 情報保障

1 はじめに

高等教育機関での障害等のある学生（以下、障害学生）への修学支援として、授業などの情報保障が実施される機会も増えてきた。情報保障では、確実に情報を伝える必要があるが、そのためには、情報の量、内容、提示方法が適切であることが重要となる。

情報保障の方法は様々で、最近では支援機器や支援技術を活用する方法が加わってバリエーションの広がりを見せている。本稿で取り上げる音声認識技術については、近年、大学等の修学支援現場で活用されるケースも増えているが、誤認識にいかに対策するかがポイントとなっている。

そこで本稿では、音声認識技術を有効に活用するための参考となるよう、情報の提示要領と情報理解の関係を確認するために行った実証実験の結果を紹介する。具体的には、音

声認識ログと動画の情報を段階的に提示し、被験者に対するアンケート形式の調査で情報の理解度を検証した。以下、情報保障手段としての音声認識の特徴を概観した後、調査と結果分析について説明する。

2 情報保障手段としての音声認識技術

情報保障とは、障害等により情報の取得が難しい場合に、代替手段等により情報を提供することを言い、障害の種類や程度によって、必要とされる情報保障の手段は異なる。高等教育機関で実施される情報保障として、表1に、日本学生支援機構の「障害のある学生の修学支援に関する実態調査」（以下、実態調査）の手引きに記載されている支援の説明をまとめた。

表1のうち、「点訳」「墨訳」「教材のテキストデータ化」「教材の拡大」「リーディ

「読み上げソフトの利用」
「講義に関する配慮」は、主に視覚情報の取得が難しい場合に有効と考えられる支援方法である。対して、「手話通訳」「ノートテイク」「パソコンテイク」「ビデオ字幕付け/文字起こし」「注意事項等文書伝達」「FM補聴器/マイク使用」「音声認識ソフト使用」は、主に聴覚情報の取得が困難な場合の情報保障手段として挙げられる。

聞こえにくい場合は、音声情報を聞き取りやすい環境を整えるか、音声情報を視覚情報に変換する方法を検討することになる。視覚情報への変換による支援のうち「手話通訳」

「ノートテイク」「パソコンテイク」はサポーターによる通訳である。「音声認識ソフト使用」はこれらの方法に加えて情報保障を実施する際のツールとして活用される場面が増加していると思われる（岡田他 2022）。音声認識の活用方法として、サポーターが聞き取った音声を端的に言い直すリスピークが挙げられる（Yamamoto 2021）が、これはサポーターが要点を伝えようとする点で通訳に近い方法といえる。通訳を介さず、音声を直接集音して音声認識する方法は、人の手を介さない手軽さがある一方、誤認識への対策も必要となる。

表1 高等教育機関における障害等のある学生へ授業の情報保障の例（概要）

支援方法	概要
点訳	教材、配布資料などを点字に訳すこと
墨訳	点字を墨字、活字に訳すこと
教材のテキストデータ化	教材、配布資料等をテキストデータにすること
教材の拡大	教材、配布資料等を拡大読書器でモニターに拡大表示したり、大きな文字で印刷したりすること
リーディングサービス	資料や配布資料などを音声で読み上げる（文字を音声に訳す）こと
手話通訳（触手話を含む）	講義内容や周りの様子を手話で伝えること
ノートテイク	講義内容や周りの様子を筆記し、文字で伝えること
パソコンテイク	講義内容や周りの様子をパソコンに入力し、文字で伝えること
ビデオ教材字幕付け 文字起こし	教材として使用される映像メディアの台詞やナレーションをテキスト化し、字幕として挿入する、または紙面に記すこと
注意事項等文書伝達	通常は口頭で受験者に伝達する注意事項を文書にして配布あるいは板書すること
FM 補聴器 / マイク使用	講義者の声を、専用のマイクを通じて補聴器を装着した学生に伝える支援技術の活用。またはその貸出などを行うこと
読み上げソフト 音声認識ソフト使用	電子データを音声に変換して読み上げたり、音声データを電子データに変換するソフトの活用。またはその貸出等を行うこと
講義に関する配慮	講義内容の録音、板書の撮影等を許可すること

（日本学生支援機構「実態調査の手引き」を一部抜粋し筆者が作成）

音声認識は、技術的には集音したすべての音声情報を認識しようとする点で、要約したり意識する通訳とは異なる支援ツールと言える。有効な情報保障とするためには、内容理解に適うだけの情報の量と質を担保する必要があるが、音声認識の活用には、誤認識への対策が必要となる。例えば、音声認識を行う環境を調整することで認識率の向上を図ったり（岡田他 202）、認識結果の中から誤認識の修正を行う（小畑 2021）など工夫がされているところである。音声認識技術をうまく使いこなす支援者によるリスピークは、誤認識対策としても有効な方法である。

音声認識による誤認識の程度によっては、情報の把握が困難になったり、必要な情報量の確保が難しくなることもある。そのため音声認識率の向上や誤認識の修正などの対策を充実させる必要もあるが、一方で、授業情報には音声認識ログ以外にも、テキストや資料、動画や講義室の様子などがあり、複数の情報源を組み合わせることで授業内容を理解する環境にもある。そこで、本稿では音声認識結果に加えて動画の情報を段階的に提示することで、内容の理解度がどのように変化するかを検証し、情報提示の要領と情報の理解度の関係を確認することとした。

3 検証調査の実施要領

本調査では、情報の提示要領と理解度の関係性を確認するために、段階的に情報の提示の仕方を変えて被験者に確認してもらい、アンケート形式で調査を行い、情報の理解度を確認した。具体的には、二種類の動画を音声認識した結果として、誤認識を修正せずに残したログと、動画（音声なし）、動画（音声あり）とを段階的に組み合わせて理解度を確認した。

検証の目的は、情報の不足や誤認識が、どの程度、情報の理解に影響するかを確認するためである。仮説として①情報量が少なけれ

ば理解度は担保されないことと、②誤認識が理解度を妨げることが予測されたが、これらを具体的に検証することとした。

以下、具体的な検証要領を説明する。

2.1 動画教材

動画はタイプの違うものを二つ用意した。いずれも大学の広報のために作成された動画で、同系統の専門分野の内容を概説した内容であるが、発話や説明要領に違いがある。表 2-1 は、それぞれの動画の特徴を簡単にまとめたものである。

表 2-1. 動画の特徴

動画	特徴
A	比較的滑舌がよい 台本を読む様な説明が多い キーワードがテロップ表示される
B	比較的抑揚がある 感嘆詞や助詞が多い 固有名詞やキーワードを手書きで表示

表 2-2. 使用した音声認識ログの概要

動画	文節数	認識率
A	121	86.77 % 音声認識：Livetalk
B	229	68.122 % 音声認識：Office365

認識率は、文節ごとの認識状況を確認し動画別に算出した。表 2-2 は、検証に使用した音声認識ログの概要である。検証では、各動画で機材や音声認識システムの組み合わせを変えて何度か試行したなかで、平均値に近いサンプルを活用した。算出した認識率は、検証中は伏せ、検証終了時に被験者に伝えた。

2.2 検証方法

各動画の音声認識ログと動画の情報を、三

段階に分けて被験者に提示し、アンケートへの回答を依頼した。なお、リアルタイムでの授業を想定し、検証は下記の要領で一人の被験者につき一度きり実施することとした。被験者には事前に実施要領を説明してから検証を開始した。

- ▶ 検証Ⅰ) 各動画の音声認識ログを、誤認識を修正しない状態のままプリントアウトで被験者に提示した。被験者には、文字情報を読んで、誤認識を思われる箇所をチェックし、正しいと予想する単語や文節を予想して記入するか、予測が不可能な場合は「？」を記載してもらった。
- ▶ 検証Ⅱ) 検証Ⅰの要領に、無音の動画を観る作業を加えた。音声認識ログは検証Ⅰと同様に、プリントアウトで提示し、被験者には同時に無音の動画を確認しながら、誤認識がありそうな箇所をチェックし、正しいと予想される語を記載するか、予測が不可能なものには「？」を記載してもらった。
- ▶ 検証Ⅲ) 被験者に音声付きの動画を観てもらい、それぞれの音声認識ログの誤認識箇所と正しい情報の確認を行った。

2.3 被験者

本検証には、学生特別支援室の学生スタッフ10名に、被験者として協力いただいた。学生スタッフは学部・学年ともに様々で、支援経験にも幅がある。今回は、音声認識や補聴システムに関する基礎的な情報を確認したうえで、事前に誤認識を見つけるための簡単な練習をして検証に臨んでもらった。

上記検証Ⅰ～Ⅲに係るアンケート回答に要した時間はスタッフ間で差があるものの、概ね1時間程度である。

3 検証結果と考察

被験者10名のアンケート調査結果の概要を表3-1にまとめた。表3-1では、被験

者が誤認識であると考えた箇所数（誤認識指摘数）と、そのうち被験者が予想した語が正解であった数（正解数）、誤認識と指摘するものの正解の単語を記載していなかったり「？」としており正解が予測できなかったと思われる箇所数（予測不能数）の平均値を示す。表中の正解率は、「正解数」を「誤認識指摘数」で割った値である。

なお、各動画の音声認識結果のうち、誤認識に当たる箇所は、動画Aは17か所、動画Bは58か所である。

表3-1. 誤認識指摘結果の概要（平均値）

動画		検証Ⅰ	検証Ⅱ
A : 誤認識数 17	誤認識 指摘数	14.1	13.1
	予測不能数	3.8	1.2
	正解数	8.3	10.6
	正解率	58.27	80.48
B : 誤認識数 58	誤認識 指摘数	18.8	15.8
	予測不能数	9.4	4.6
	正解数	2.3	5.9
	正解率	12.07	41.82

3.1 検証Ⅰと検証Ⅱの理解度

いずれの動画でも、「予測不能数」は検証Ⅰより検証Ⅱで減少しており、また「正解数」「正解率」ともに検証Ⅰに比べて検証Ⅱで上昇している。このことから、検証Ⅰより検証Ⅱの方が、内容の理解度が向上していることがうかがえる。

検証Ⅰでは、音声認識ログのみを提示したが、検証Ⅱでは音声認識ログに加えて無音の動画の情報を加えた。動画Aと動画Bでは動画の撮影方法が異なるものの、いずれもキーワードや固有名詞がキャプション等で提示されている部分や、作業の様子や話の流れを図

示する部分があり、これらの視覚情報が直接的に情報理解を促したり、音声認識ログの誤り部分の補正に役立ったと予測される。

一方で、いずれの動画でも誤認識の指摘数が検証Ⅱで減少している。この原因については不確かなものの、情報量が増えたことで誤認識ではなく正しい情報であると判断した可能性が考えられる。ネガティブな解釈をすれば、検証を続けて行うことで被験者が情報の内容に慣れてきた可能性や、立て続けの情報量に疲労してきた可能性などが挙げられる。

なお検証Ⅲでは、音声付きの動画視聴により、検証Ⅱまでで予想した内容の正答を被験者自身が確認した。被験者には検証Ⅲが終了した時点で、それぞれの動画の認識率がどれくらいかの予想や実証実験の感想を聞いているが、この点については別の機会に紹介したい。

3.2 被験者ごとの理解度

被験者ごとの結果として、表3-2に、誤認識を指摘したものの正しい語が予想できなかった箇所の数「予想不能数」を、表3-3に、誤認識と指摘し正しいとして予想した語が正解であった「正解率」を示す。

表3-2. 被験者毎の「予測不能数(個)」

被験者	動画A		動画B	
	検証Ⅰ	検証Ⅱ	検証Ⅰ	検証Ⅱ
1	0	0	10	7
2	3	3	11	11
3	6	1	20	1
4	6	1	5	2
5	4	2	6	4
6	3	1	11	3
7	5	3	5	4
8	3	1	11	3
9	4	0	9	8
10	7	1	7	1

表3-3. 被験者ごとの「正解率(%)」

被験者	動画A		動画B	
	検証Ⅰ	検証Ⅱ	検証Ⅰ	検証Ⅱ
1	80.00	93.75	19.23	25.93
2	75.00	75.00	08.70	23.81
3	46.67	83.33	07.14	75.00
4	35.71	87.50	16.67	38.89
5	53.85	71.43	14.29	30.77
6	62.50	75.00	10.53	36.36
7	66.67	78.57	11.11	40.74
8	68.75	86.67	00.00	33.33
9	57.14	78.57	17.65	46.67
10	36.36	75.00	15.38	66.67

表3-2中に、検証Ⅰに比べ検証Ⅱで、誤認識と思われながらも正しい語を予測できなかった数が減少している部分を網掛けして記した。被験者の多くで、予測不能数が減っており、動画Aでは10名中8名、動画Bでは9名となっている。

表3-3では、ほとんどの被験者で正解率が検証2で上昇している。10名中9名が上昇、同値が1名(被験者2/動画A)で、被験者によって値に開きがあるものの、最大で67.86(被験者3/動画B)の伸びが確認された。

3.3 動画ごとの理解度

動画ごと検証結果を確認してみると、表3-2では、動画Aに比べ、動画Bでは、被験者ごとまた検証ごとの「予測不能数」のばらつきが大きい印象である。音声認識ログ内での誤認識箇所は動画Bの方が多いが、指摘数は大きく届かない。

表3-3の「正解率」を見ると、数値は動画AとBでは大きく開きがあり、全体的に動画Aに対して動画Bの数値は低いことが分かる。例えば動画Aは、検証Ⅰの段階で最低でも30を超えているのに対し、動画Bの最低

値は0である。検証Ⅱではいずれの動画でも数値は上昇しているものの、動画Aの正解率が最高で93.75であるのに対し、動画Bの最高値は75に留まっている。

被験者によって数値に開きはあるものの、動画Aは誤認識を指摘しやすく、また正しい語を予測しやすい傾向があるが、動画Bは誤認識かどうかの判断が難しかった可能性が指摘できる。

3.4 誤認識の指摘例

音声認識ログのうち、どの語句が誤認識だと指摘されたのか、また語句の正誤はどのように判断されたのかについて検討するため、特徴的な例を表3-4に挙げる。

表中の「指摘語句」は被験者が誤認識として指摘した語句、「正解語句」は動画作成側が本来伝えようとした語句、「正解数 / 指摘数」は、誤認識と指摘した被験者数のうち正解語句を予測できた人数を示している。

例1と例2は、多くの被験者が誤認識と予測しながら検証Ⅰでは、正しいと思われる情

報を予想しにくかった例である。例1は専門用語であること、例2は文脈とかけ離れた誤認識であったために「正解語句」の予想が難しかったと予測できる。いずれも検証Ⅱでの正解数が上がっていることから、動画のテロップ等の視覚情報が「正解語句」を導き出すヒントとなっていたことがうかがえる。例3と例4も同様のことが指摘できるが、検証Ⅱの段階で被験者全員が正解語句にたどり着くことができた例として挙げておきたい。

例5は、多くの被験者が誤認識と指摘しながら、検証Ⅱでも「正解語句」を想像しにくかった例である。この例では「正解語句」を“菌子”と予想した被験者が5名おり、同音異義語の理解の難しさが露呈した結果となっている。

今回の検証では、例6のように、正解である語を誤認識と指摘する例もいくつか確認された。このような誤指摘は、何が正しい情報かの確信が無い状況では、起こりうる現象である。

表3-4. 指摘語句の例

例	指摘語句	正解語句	正解数 / 指摘数	
			検証Ⅰ	検証Ⅱ
1	TheRoom	フザリウム	0/9	9/9
2	3月	畜産学	0/9	6/8
3	本誌	孢子回収	7/10	10/10
4	今米ションマーティ	イノベーションマーケティング	2/10	10/10
5	禁止	菌糸	3/10	2/9
6	気温	気温	0/1	0/1

4 考察

本検証の結果からは、被験者や動画により値に差はあるものの、検証Ⅰと検証Ⅱを比較すると、概ね予測不能数が減少し、「正解率」が増加するという傾向が見て取れた。

被験者は、視覚情報や前後の文脈などから、

誤認識されたとと思われる語句を正しいものに修正して理解しようとしたが、この時、予測した単語が正しければ、音声認識ログの認識率に加えて情報の正確性が増すことになる。この点を確認するために、被験者が補正した情報を加えた情報の正解率を「補正正解率」

として求めたい。「補正正解率 (%) = ((正解数 / 被験者数) + (音声認識ログの正答認識文節数)) / 音声認識ログの総文節数」という式により算出した数値を表 4 に記す。

表 4. 補正正解率 (%) : 有効数字を表示

動画	認識率 (%)	補正正解率 (%)	
		検証 I	検証 II
A	86.77	92.89	94.71
B	68.12	69.12	70.56

※「補正正解率 (%) = ((正解数 / 被験者数 + 音声認識ログの正答認識文節数)) / 音声認識ログの総文節数」として算出

補正正解率は、誤認識の予測を含めた内容理解の可能性を示す数値と言い換えることができる。音声認識ログによっては正しい情報の予測が容易なこともあり、情報をカバーできる可能性がある。検証 I から検証 II で数値の伸びが確認されているとおり、動画の情報が加わればこの点はさらに強化されている。

動画ごとの補正正解率をみると、動画 A のほうが優位であるだけでなく、音声認識ログの認識率 (86.77%) に比する補正正解率の値が検証 I では 92.89% と 6 ポイント、検証 II では 94.71% で 8 ポイントほど情報が補正されている。これは、動画 B での伸び率が検証 I では 1 ポイント程度、検証 II では 2 ポイント強に留まっていることと対照的である。動画 A は、そもそも音声認識ログの認識率が高いことに加えて、テロップ等の視覚情報が効果的で、これらの要因が高い補正正解率に結びついたと考えられる。

今回の検証では、音声認識ログの情報を補填する形で無音の動画情報を加える方法を使った。検証結果からは、基本的には情報量が増えるほど理解度は向上する傾向にあると指摘できる。

一方で、情報量が多いほど理解が促される

かという疑問が残る。例えば、動画 B の検証 II では、無音の動画の情報が加わっていながら、正しい内容を推測することが難しい状況が見て取れた。誤った情報が多出すれば内容理解が妨げられてしまうことになる。

正しい情報が多くあれば理解が促進される傾向にあるが、情報は適度な量で、適当な方法で提示することも重要である。誤認識の指摘数は検証 II で減少していたが、この背景には、多くの情報を一度に処理することの難しさ、特に複数の情報源を一度に視認し統合して理解することの難しさも可能性として挙げられる。

4 結論

本稿では、情報の提示要領と情報理解の相関関係を確認するための検証を行った。これまでの議論を踏まえ、音声認識を活用した情報保障の充実のために必要なことを整理しておきたい。

音声認識技術は、集音した音声を文字に変換するものである。この特性は、活用の気軽さとともに誤認識のリスクを伴っている。誤認識への対策を採りながら、より正しい情報を、適度な量と方法で提示する必要がある。本稿では、音声認識ログに動画の情報を加えていったが、音声情報の保障を充実させる方法は、音声認識の活用といった一つの手段を駆使するだけでなく、複数の視覚情報を組み合わせる方法にも活路はある。情報の読み取り結果の分析では、視覚情報や前後の文脈、一般的な知識などから誤認識されたログを適宜修正できている箇所もあれば、専門用語や同音異義語の誤認識に苦戦したり、正しい情報を誤認識であると指摘するケースも見られた。このことから、専門用語やキーワードは、確実に伝わるよう、文字を視認できるよう提示する工夫が有効であると指摘できる。

複数の視覚情報を組み合わせる場合は、視線の集約を図る必要がある。手話通訳やノー

トテイク等の情報保障に関するテキストでは、音声情報を視覚情報に変換して伝える際には、確認する対象をなるべく近い位置に配置し、一度に視認しやすくするようレイアウトを工夫することが指摘されている（日本聴覚障害学生高等教育支援ネットワーク情報保障評価事業グループ 2007:38, 「要約筆記者養成テキスト」作成委員会 2015:39）。今回の検証では、紙面で音声認識ログを配布し、被験者は検証Ⅱにおいて、紙面と動画画面の両方を同時に目で追う必要があったが、この場面で被験者たちは結果的に、日常的に支援現場で起こりうる困難に陥っていた可能性がある。

5 おわりに

本稿は、近年多くの高等教育機関にて活用されている音声認識技術によるサポートを想定し、情報の提示の仕方と理解度の関係性を読み解く試みを行い、音声認識技術をうまく活用するために、誤認識の対策と同時に、他の情報源も活用しながら、正確な情報を適切に提示することの重要性を改めて指摘した。

検証作業では、学生特別支援室の学生スタッフに協力を仰いだ。手元にある音声認識ログのどこが正しくどこが誤っているのかが判然としない中で、限られた文字や動画の視覚情報を頼りに正しい内容を予測しながら理解しようとする、今回の検証で被験者たちが置かれた状況は、聞こえにくい学生達が日常的に修学の現場で直面している現象でもあった。

技術の進歩は目覚ましく、より安価で便利な技術を活用する機会も増えている。支援技術や支援機器を現実的かつ有効な支援の手段に落とし込むためには、手軽で確実な運用と、授業情報が確実に学生たちの手元にとどくことの両面での工夫が必要となる。

付記／謝辞

本稿で行った検証と結果の分析は、学生特別支援室学生スタッフ森岡龍治さん（農学部

4年）と検討して実施したものです。森岡さんとは、特に機材を活用した情報保障について、現実的に取りうる方法を検証したり、情報保障上の課題を確認したりしてきました。今回の調査研究もその一環として実施されたものです。

また、アンケート調査には、本年度の学生スタッフの面々が協力してくれました。本稿を支える調査及びその結果は、学生特別支援室の学生スタッフの活躍により得られたことをここに改めて紹介します。

支援現場での実感と調査分析による客観的な視点とを併せて、現実的な対策や課題を検討できる機会になったなと思います。支援・教育・研究が連動することで、今後の大学における情報保障がいつそう充実することを期待します。

(学生支援センター 学生特別支援室
准教授)
(学生支援センター 学生特別支援室
学生スタッフ)
(教育学部 准教授
・学生支援センター 学生特別支援室室長)

【参考文献】

- (1) 日本学生支援機構「障害のある学生の修学支援に関する実態調査」
https://www.jasso.go.jp/statistics/gakusei_shogai_syugaku/index.html
- (2) 日本学生支援機構「令和4年度(2022年度)大学、短期大学及び高等専門学校における障害のある学生の修学支援に関する実態調査 調査の手引き」
https://www.jasso.go.jp/statistics/gakusei_shogai_syugaku/_icsFiles/afieldfile/2022/08/26/2022_tebiki.pdf
- (3) 岡田菜穂子・森岡龍治・藤本菜摘・源川礼菜・川口真菜・田村鈴歌・赤尾侑果・川

相歩未・野上明里・浅田麻莉絵・近藤和
奏・江崎渚月・増田悠希・須藤邦彦2022
「音声認識技術と補聴システムの活用によ
る授業情報保障の向上のための実証実験」
『大学教育』第19号, 42-49.

(4) Mikio Yamamoto ・ Akiko Sakamoto ・
Nahoko Okada, 2021, Practice and
automation of remote real-time
captioning support for students with
disabilities using speech recognition
technology, 2021 10th International
Congress on Advanced Applied
Informatics, 934-935.

(5) 小畑千尋, 2021 「オンラインによる「音
楽科教育法（初等）」での情報保障の実際
と課題—聴覚障害学生への支援を通して—」
『宮城教育大学 情報活用能力育成機構研
究紀要』第1号, 1-8.

(6) 日本聴覚障害学生高等教育支援ネットワ
ーク情報保障評価事業グループ, 2007 『大
学ノートテイク支援ハンドブック』人間社

(7) 「要約筆記者養成テキスト」作成委員会,
2015 『要約筆記者養成テキスト上』（初版
第6刷）全日本難聴者・中途失調者団体連
合会・全国要約筆記問題研究会