

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

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

要旨

障害等のある学生へ情報保障の方法として音声認識技術の活用が進んでいるが、より有効に活用するためには認識率の担保や運用方法の工夫が必要となる。本稿では、音声認識技術と補聴システムとを組み合わせた情報保障の質の向上のために、端末や機器の使用方法を工夫する余地がどの程度あり得るのかについて、実証実験結果をもとに分析する。実証実験からは、同じ動画教材でも音量の調節や、機材の組み合わせにより認識率に差異があることが明らかとなった。一方で、動画教材の音源の鮮明度等によっては、技術や機器の使用方法による認識率に大きな開きは見られなかった。このことから、情報保障機材等の工夫がある程度有効であるものの、発話や教材作成時の雑音への配慮が必要であると言える。

キーワード

高等教育機関，障害学生，修学支援，音声認識，補聴システム，情報保障

1 はじめに

障害等のある学生の修学支援上の情報保障の方法は様々であるが、近年では支援機器を活用する機会も多く見られる様になった。日本学生支援機構では平成18年度から「障害の

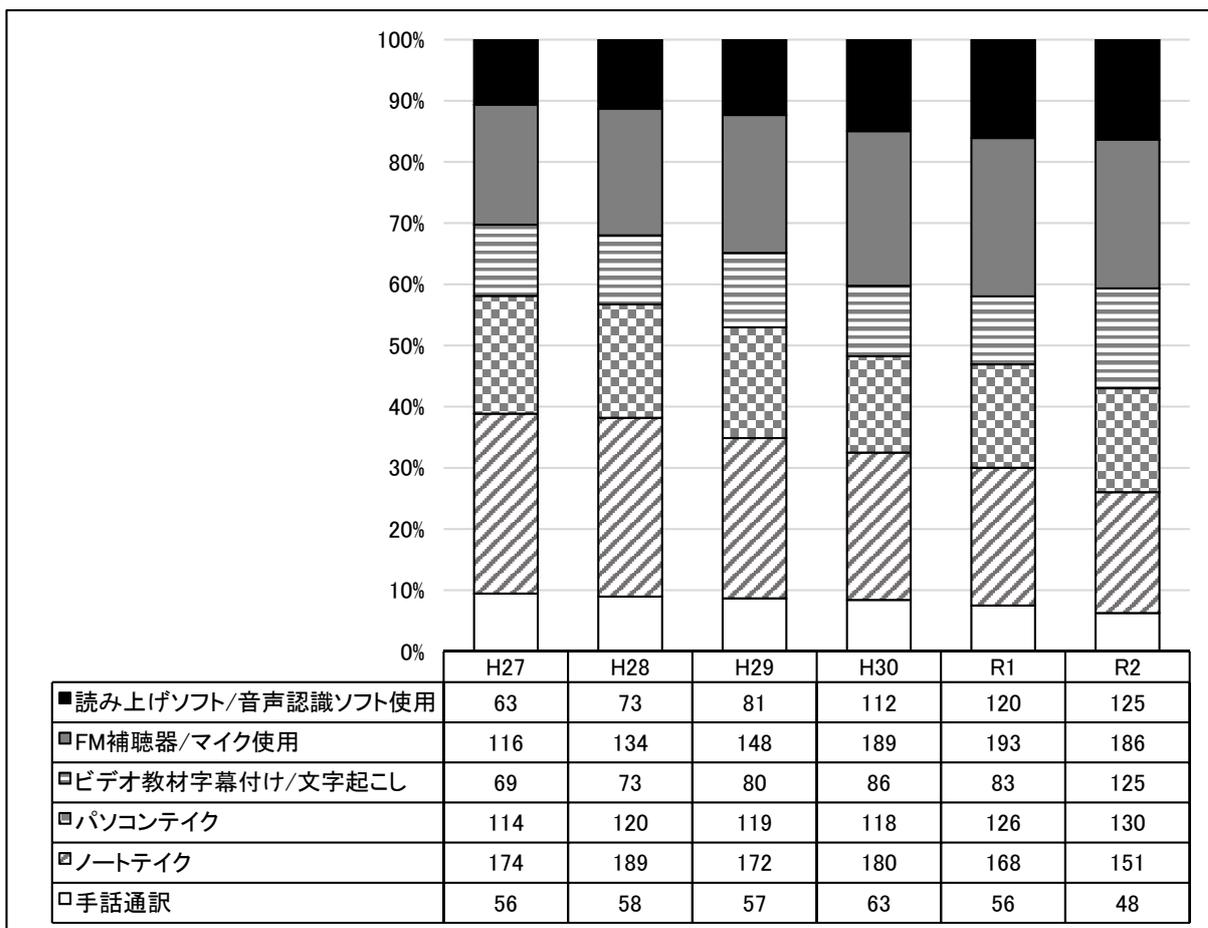
ある学生の修学支援に関する実態調査」（以下、実態調査）を実施しており、全国の高等教育機関を対象に、障害等のある学生の在籍状況や修学支援の実施状況に関する調査結果を公開している。

グラフ1は、実態調査の報告から、授業支援として情報保障にかかわる項目を抜粋してまとめたものである。グラフ1によると、障害等のある学生への授業支援実施状況として「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」といったいわば支援者を介して行う支援に加え、「FM補聴器・マイク使用」「読み上げソフト・音声認識ソフト使用」といった、支援機器や技術による支援を実施している教育機関が増加していることが分かる。

実態調査に障害等のある学生への授業支援方法として「音声認識ソフト使用」の項目が加わった平成27年度から令和2年度までの調査結果を見ると、この期間内に「手話通訳」や「ノートテイク」が微減し「パソコンテイク」

が微増しているのに対し、「FM補聴器・マイク使用」「読み上げソフト・音声認識ソフト使用」を実施している教育機関数が増加しており、年を追うごとに実施校数が伸びを見せている。結果として、平成27年度は情報保障に関わる支援実施方法として「ノートテイク」を採用している大学等が最も多かったが、令和2年度は「FM補聴器・マイクの使用」が最も高い値となっている。かといって「ノートテイク」等の人的支援の実実施校数が急激に減っているわけではないことから、支援機器が取って代わったというよりは、これまでの支援者によるサポートに加えて、新たに支援機器や技術を導入・併用している教育機関が増えているという見方が妥当であろう。

グラフ1．障害等のある学生への授業支援実施状況（実施校数）



（日本学生支援機構「実態調査」より作成）

この背景には、支援技術の進歩により、支援のための人材の安定的な確保という課題（岡田 2018，坂本 2018 等）への対応が容易になってきたことや、感染症対策として遠隔授業の情報保障の必要性が高まったこと（岡田 2021）等があると考えられる。遠隔授業の実施により支援ニーズの変化が起こっている様子は、グラフ1「ビデオ教材字幕付け／文字起こし」の値が令和2年度に急増していることから予測できる。

高等教育機関での授業情報保障のための音声認識技術の活用については、これまで教育現場での実践報告や、有効性の検証等が行われてきた。有効な情報保障とするためには、授業理解に適うだけの情報の正確さを担保する必要があるが、音声認識については認識率の向上や誤変換への対応が要となる。そこで、講師の音声をサポーターが聞き取って端的にリスピークする方法や（Yamamoto 2021）、誤認識をサポーターが修正する方法（小畑 2021）などが採用されている。音声認識技術の向上には目覚ましいものがあるが、それでも認識率は完璧ではない。誰がどのタイミングで授業情報を音声認識にかけるのか、誤認識への対応をどのように行うのかといった、運用の問題上の工夫により、より効率的かつ効果的な情報保障の実現が求められている。

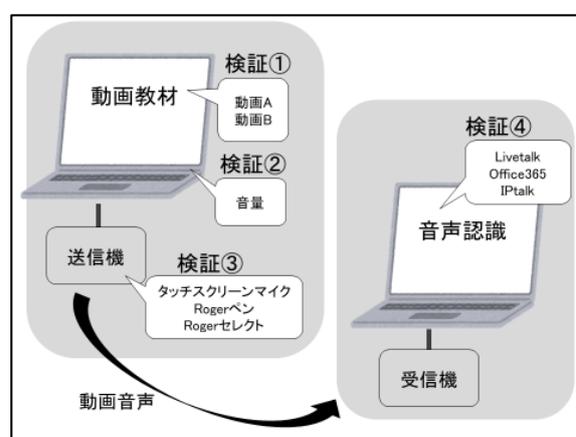
認識率の向上のためには、音声の送信時と、受信時に工夫の余地があるが、これまでの高等教育機関における音声認識活用に関する先行研究では、情報の受信時の調整の可能性についてはあまり触れられていない。本稿では、情報の送信時と受信時の双方に着目し、使用機材を調整し、認識率の高い環境の整備がどこまで可能かを検証したい。具体的には、音声認識と補聴システムの組み合わせと、動画送信時の音量の違いによる認識率を測り、比較検証結果を分析する。その結果から情報送信時の工夫について言及したい。

2 検証方法

今回の実証実験では、複数の音声認識システムと補聴システムを組み合わせ、異なるタイプの動画教材の音声認識率を比較した。

具体的には、図1に示すように①A・Bの動画教材を使用して（表1）、②音量を調整しながら、③補聴システムの送信機（表2）で音声を送信し、④音声認識システム（表3）を搭載したパソコンで受信し、これらの組み合わせで認識率を測った。

図1. 検証方法の概要



2.1 動画教材

動画はタイプの違うものを二つ採用した。いずれも大学の授業等で活用されている動画教材で、同系統の専門分野に関する内容であるが、発話や説明要領に違いがある。表1は、それぞれの動画教材の特徴を簡単にまとめたものである。

表1. 実証実験で使用した動画教材

動画	文節数	特徴
A	121	比較的滑舌がよい 台本を読む様な説明が多い
B	229	比較的抑揚がある 感嘆詞や助詞が多い

2.2 補聴システム

補聴システムとしては、PHONAC社の

Roger を活用した。Roger はデジタル補聴援助システムで、クリアな音声を補聴器や受信機に届けることができる。様々な機器が展開されており、機器を組み合わせたり、モードを切り替えることで、講義やプレゼン、グループディスカッション、インタビュー等、用途に合わせて活用することができる（山口大学学生特別支援室 2021b）。

音声認識システムによっては、音声認識ソフトを搭載した端末上で動画を再生することで音声認識する機能を有するものもある。今回の実証実験では、補聴システムによる認識率の差異を図ることが目的の一つであるため、敢えてこの機能を使わず、補聴システムを介して動画音声を送信し、その音声を音声認識搭載端末にて受信する方法を採った。

表 2. 実証実験で使用した Roger 送信機

Roger 送信機	実験で使用
タッチスクリーンマイク	○
Roger ペン	○
Roger セレクト	○
マルチメディアハブ	

2.3 音声認識システム

本実証実験では、以下の 3 つの音声認識システムを取り上げた。いずれも大学の授業等での情報保障のために活用されているものであり、比較的入手しやすい製品である（山口大学学生特別支援室 2021a）。

表 3. 実証実験で使用した音声認識システム

製品名	音声認識エンジン
富士通 Livetalk	Amivoice ローカル
MSoffice365	ディクテーション
IPtalk	Google 音声認識機能

2.4 動画再生の音量

本実証実験では、動画を再生する端末上で音量を調整することで、音量による認識率に違いがあるかについても着目した。なお、使用した端末は全てマイクロソフト社の Surface である。送信時の端末の音量を 20、40、60、80 として認識率を調べた。

2.5 認識率の算出方法

音声認識の認識率については、主に文字単位での正誤を確認する CER(Character Error Rate) と、単語単位で正誤を確認する WER(Words Error Rate)があるが、本研究では WER を採用し、文節ごとに認識状況を確認し認識率を算出した。

3 実証実験結果

実証実験結果を、動画、音量、補聴システム、音声認識システムごとに見てみる。

3.1 動画による認識率の差

動画ごとの認識率は、表 4 の通りである。平均認識率は、動画 A と動画 B では約 20 % の開きがある。また動画 A は最高認識率と最低認識率の差が 6 % 弱であるのに対して、動画 B は 90 % となっており、動画 A に対して動画 B は単語ごとの認識率にばらつきがあり、全体として認識率が安定していないことが指摘できる。

表 4. 動画別認識率 (%)

	平均	最高	最低
動画 A	87.328	89.256	83.471
動画 B	68.506	90.830	0.437

3.2 音量による認識率の差

同一のパソコン機種 of 端末から音声を送信する際に音量を調整して音声認識率を測った結果として、表 5 に動画 A、動画 B それぞれで認識率を示した。

表5は、音声認識システムや補聴システムの組み合わせを変えて得た全サンプルのうち、認識率上位25位までのものについて、動画ごとに音量情報を併記したものである。

表5. 音量を調節した場合の認識率 (%)
(上位25位までの値で作成)

動画 A		動画 B	
認識率	音量	認識率	音量
89.256	40	90.830	40
89.256	40	90.830	40
89.256	40	90.393	40
89.256	20	89.520	40
89.256	20	89.520	40
88.430	40	87.773	40
88.430	40	87.336	40
88.430	40	87.336	40
88.430	20	87.336	40
87.603	40	86.900	40
87.603	40	86.463	40
87.603	40	85.153	40
87.603	40	84.279	40
86.777	40	81.223	40
86.777	40	78.603	60
86.777	20	77.729	40
86.777	20	74.672	20
86.777	20	72.052	80
86.777	20	71.616	20
86.777	20	68.996	20
86.777	20	68.122	20
86.777	20	67.249	20
86.777	20	66.812	20
85.950	20	65.502	80
85.124	20	64.192	80

表5によると、動画A、Bとともに、音量40が上位に多く見られる。動画Aについては、音量40と音量20の認識率が同値のものも複数確認できるが、動画Bは、上位を音

量40が占めている。認識率と音量の関連性について言うなら、動画Aについては音量の影響は比較的少なく、動画Bについては音量によって認識率が左右されており、最適な音量は40であると指摘できる。

なお今回の実証実験では、音声認識システムによって音量の影響を受けやすいものとそうでないものがあると仮定されたが、この点を検証するためには、更に複数の教材での実証実験を行う必要があるため、今後の課題としたい。

3.3 補聴システムによる認識率の差

補聴システム3種類の認識率を表6にまとめた。なお表6の認識率は、各補聴システムのうち、音声認識ソフトや音量の組み合わせで平均値を記載している。

表6. 補聴システム別認識率 (%)

	タッチスクリーンマイク	Rogerペン	Rogerセレクト
動画 A	87.33%	87.33%	87.33%
動画 B	72.77%	66.02%	66.73%

動画Aでは、全ての補聴システムで認識率同値87.33%であった。動画Bでは、RogerペンとRogerセレクトが66%代と同程度の認識率で、それより約6%高い認識率がタッチスクリーンマイクである。動画Aと比較すると、動画Bについては補聴システムの種類による認識率の幅が見られ、タッチスクリーンマイクとの相性が最も良いという結果になった。

3.4 音声認識システムによる認識率の差

音声認識システム3種類について、動画ごとの平均認識率を表7に示す。

表7によると、動画Aでは、音声認識システムによらず、概ね平均認識率が良好で、音

声認識システムによる認識率にさほど差はない。一方、動画 B では動画 A に対して平均認識率が低く、また音声認識システムによっては認識率が極端に低いことが分かる。

表 7. 音声認識システム別平均認識率 (%)

	Livetalk	IPtalk	Office365
動画 A	87.33	87.47	87.28
動画 B	82.10	45.01	81.90

この点についてさらに検証するため、表 8 に認識率の良いサンプル上位 25 位までが、どの音声認識システムを使用したものかをまとめた。表 7 においても、動画 A の認識率は安定しており、認識率に大きな開きはない。対して、動画 B では Office365 が上位を占めていて、IPtalk はあまり目立たない。しかも認識率の差が大きく、認識率の最高値が

Office365 で 90 % を超えるものの、16 位からは 8 割を切り 21 位からは 7 割を下回っている。IPtalk では最高値でも 8 割に達していないことから、動画 B と IPtalk との相性は悪いと言える。

表 8. 音声認識システムによる認識率
(上位 25 位までのサンプルで作成)

動画 A		動画 B	
認識率	音声認識	認識率	音声認識
89.256	IPtalk	90.830	365
89.256	IPtalk	90.830	365
89.256	365	90.393	365
89.256	365	89.520	365
89.256	365	89.520	365
88.430	Livetalk	87.773	365
88.430	365	87.336	365
88.430	365	87.336	365
88.430	365	87.336	Livetalk
87.603	365	86.900	365
87.603	365	86.463	365

87.603	365	85.590	365
87.603	365	84.279	365
86.777	Livetalk	81.223	Livetalk
86.777	Livetalk	78.603	IPtalk
86.777	IPtalk	77.729	Livetalk
86.777	IPtalk	74.672	365
86.777	IPtalk	72.052	IPtalk
86.777	365	71.616	365
86.777	365	68.996	365
86.777	365	68.122	365
86.777	365	67.249	365
86.777	365	66.812	365
85.950	IPtalk	65.502	IPtalk
85.124	365	64.192	IPtalk

※ 365 : Office365

4 考察

本実証実験では、二つの動画教材を材料に、補聴システム、音声認識システム、音量による音声認識率を測った。

動画 A では、音量、補聴システム、音声認識システムそれぞれの実験の結果、音声認識率に開きはさほど見られなかった。動画 A については、機材やシステムとの相性という点では、環境による認識率に大差が無いという結果となる。敢えて言えば、動画 A の音量別の認識率では音量 40 が上位に多いことから、音量適性は 40 であると指摘できる。

動画 B では、音量、補聴システム、音声認識システムで、認識率の差が大きく、最大で 90 % もの差が生じていた。使用する機材や音声認識システムとの相性が生じやすいという特徴が見て取れた。送信機の機器別ではタッチスクリーンマイクの認識率が安定して高く、音量別では音量 40 が圧倒的に認識率の上位を占めており、音声認識ソフト別では、Office365 が上位に多いことから、動画 B の視聴に際して、最もいい相性の組み合わせは、タッチスクリーンマイク + Office365 + 音量

40 であると言える。

5 結論

本稿では、音声認識システムと補聴システムの活用により情報保障の質をより高めるための工夫について、音声を送信する補聴システムや、音声認識システム、音量といった技術や機材の組み合わせが、どの程度効果的なのかを検証した。

実証実験では、動画によって補聴システムや音声認識システム、音量との相性が生じることから、情報保障環境の調整がある程度有効であることが明らかとなった。一方で、動画Aはどの組み合わせでも、概ね安定して高い認識率を保持していたことから、動画の特徴から音声認識率が左右されていることが推測される。つまり動画の質が担保できれば、環境に影響されず認識率を担保できる可能性がある一方、動画の質によっては機材やシステムとの相性が明確となり、これら送信・受信環境の調整が難しい場合は認識率の向上は難しいということになる。

新たな疑問は、「なぜ、動画による認識率に大きなばらつきが生じるのか」また、「その点を克服するためにはどのような工夫ができるのか」である。音声認識ソフトの特徴としては、雑音が少なく、滑舌が良いものに関しては認識率が高い傾向にあるが、今回の実証実験では、さらに助詞の認識率が悪く、それが動画Bの認識率が大きく落ち込んだ原因ではないかと推測される。

実証実験では、認識率に着目したが、誤認識した単語や文節によっても情報理解への影響が異なるケースは容易に想定される。例えば、認識率が高かったとしても、重要な単語や専門用語が誤認識されると授業内容の理解度は高まらない可能性があるし、誤変換された単語や文節でも、同音異義などであれば正しい語が推測可能なこともある。今後は、他のパターンの動画による実験を行ったり誤認

識のパターンの分析等を通じて、どのような教材や支援が分かりやすさを支えるものなのかの解明に繋がればと考える。

6 おわりに

本稿は、大学における授業の情報保障について、効率的かつ効果的な方法を模索する実証研究の一つである。今回の実証実験では、近年多くの大学で導入が進んでいる音声認識システムの活用について、機材の調整や工夫に着目した。その結果、音声認識システムや補聴システム、音量等を工夫することで、ある程度認識率を向上できる可能性が明らかになった。一方、動画の特徴によってはどのような組み合わせでもある程度高い認識率を保てることが分かった。

このことは、支援技術や支援機器を利用する側が工夫できる余地が残されているということと共に、授業情報を発信する側の工夫により、情報保障の質を向上できる可能性があることをも示唆している。本稿では、音声認識の認識率を取り上げたが、さらに視点を広げるなら、テキストや資料の提示の仕方、動画や教室の視覚情報など、授業情報の理解を促す材料は数多く存在している。

合理的配慮という言葉が使われ始めて暫くたつが、授業情報を提供する側と、受け取る側、そして情報保障に協力する立場、それぞれの工夫や歩み寄りによる対話や調整の重要性が改めて認識される結果となった。

付記／謝辞

本論に関わる実証実験は、学生特別支援室学生スタッフ森岡龍治（農学部3年）さんが中心となり実施しました。森岡さんは、実験の実施要領を工夫し、時間をかけて分析を行ってくれました。機材の組み合わせや音量の調節などは、支援に関わる学生ならではの視点で生まれたものです。

本稿を支える実証実験結果は、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。この取り組みが、今後の大学における情報保障の充実につなが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) 岡田菜穂子・小川勤・田中亜矢巳・金子博, 2018「障害のある学生のための支援者育成と課題：山口大学におけるノートテイク育成の事例から」『大学教育』第15号, 36-43.

(3) 坂本晶子, 山本幹雄, 山崎恵里, 大高下さゆり, 佐野(藤田)真理子, 吉原正治, 2018, 「大学における情報保障のための遠隔サポート実証実験の報告」『総合保健科学』第34号, 59-66.

(4) 岡田菜穂子・須藤邦彦・田中亜矢巳・柳下雅子, 2021, 「コロナ禍の障害学生修学支援」『大学教育』第18号, 38-48.

(5) 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.

(6) 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.

(7) 山口大学学生特別支援室, 2021a, 「SSR ニュースレター」10号
http://ssr.ssc.oue.yamaguchi-u.ac.jp/document/newsletter_10_202103.pdf

(8) 山口大学学生特別支援室, 2021b, 「SSR ニュースレター」11号
http://ssr.ssc.oue.yamaguchi-u.ac.jp/document/newsletter_11_202112.pdf