

## 放送大学における学習環境のアクセシビリティ： 聴覚障害者のための音声認識技術を活用した講義の字幕化

広瀬 洋子<sup>1)</sup>

### OUJ's Challenge in Improving Accessibility to Higher Education Learning Environments： Capturing Lectures Using Speech Recognition Systems for the Hearing Impaired

Yoko HIROSE

#### 要 旨

本稿の目的は、放送大学のメディア教材のアクセシビリティに関する現在の状況と、これまで継続的に行われてきたアクセシビリティ向上のための研究開発について検討し、今後の本学における聴覚障害学生への支援システム構築の一助とすることである。

2016年に施行された障害者差別解消法によって、本学は国立大学と同様に障害のある学生への合理的配慮が義務化された。本稿では、まず、現在、本学のTV、ラジオ、オンライン等で配信されているメディア教材をアクセシビリティの観点から吟味する。そして、本学において取り組んできた合理的配慮と、情報通信技術を生かしたアクセシビリティ向上のための研究開発について紹介する。次に本稿の中心的テーマである、情報保障の面で課題が残るラジオ授業に着目し、2013年から開始した京都大学との共同研究：講演・講義を対象とした自動音声認識の研究開発を紹介する。最後に、研究開発の結果、実現した複数のラジオ授業、オンライン授業の字幕付与方法について制作面から検討し、音声認識技術を活用した講義の字幕付与、それに伴う新しいコンテンツの開発にも触れながら、その成果と課題、今後の展望について議論する。

#### ABSTRACT

This paper describes the current situation and the continuous challenges faced in making educational audio materials available to hearing-impaired students. I firstly briefly discuss the legal duty of the Open University of Japan (OUJ) — in accordance with the Act of Eliminating Discrimination against People with Disabilities, enforced in 2016 — to develop information communication technologies. Secondly, I examine the present accessibility to audio educational materials concerning OUJ lectures on TV, Radio, and the Internet. I lastly look at the OUJ's radio lectures and examine the results of the experimental joint project between Kyoto University and OUJ started from 2013 that focused on a speech-recognition system for Japanese language lectures. The OUJ's audio educational materials and the digital data of text books have greatly contributed to building up the database of this system. By using this system, we succeeded in capturing several radio lectures and online course materials, thereby making the higher education learning environment more accessible to various types of students.

#### 1. はじめに

2016年度に施行された障害者差別解消法では、障害

者の社会的障壁の除去について「必要かつ合理的な配慮」を行うことが義務づけられ、高等教育においては国公立大学では義務化、私立大学は努力義務となった。放送大学は私立大学ではあるが、差別解消法施行

<sup>1)</sup> 放送大学教授（「情報」コース）

令の中で、国公立大学と同様に合理的配慮は義務づけられている<sup>1</sup>。

本学の合理的配慮としては、単位認定試験については、個々の学生の障害の特性に応じて試験時間の延長、別室受験や支援機器の利用を認めている。試験の措置として、点字受験、音声受験、代理による回答記入などの制度が講じられている。また、視覚障害の学生のために、印刷教材のテキストデータを配布するシステムが確立され、視覚障害者向けの情報サイトには、音声読み上げソフトが読みやすい形で重要な情報や点字化した印刷教材のリスト等が掲載されている<sup>2</sup>。

放送大学では2018年度に障害を自己申告した学生は811人、全体の約0.9%であるが、申告しない学生も多い。60代以上の学生数が全体の25%以上を占めていることを考えれば、視力や聴力、移動などに問題のある学生への支援のニーズは潜在的に高いと考えられる。また、本学は2018年7月現在、全国397校の高等教育機関と単位互換協定を締結しているため、この制度を活用してフレキシブルな学びを提供する本学で単位取得する障害のある学生も少なくない。よって本学の障害者への配慮と学習環境のアクセシビリティの向上は、日本の高等教育における障害学生支援にとって大きな意味をもつ。

本稿では、本学のTVやラジオ、インターネットを介して学生に届けられるメディア教材のアクセシビリティに着目し、聴覚障害者への合理的配慮として講義音声の情報保障について議論する。

## 2. 放送大学のメディア授業のアクセシビリティの現状と課題

我が国最大の遠隔高等教育機関である放送大学は、約300科目の授業がTVもしくはラジオで配信されており、さらに、その大半がインターネットでも配信され、学生はスマートフォンやタブレットなどで視聴できるようになっている。また、2016年度にはすべての学習をインターネット上の講義や課題解答で行う「オンライン授業」が開設された。以下、それぞれの授業を、合理的配慮としての情報保障とメディアのアクセシビリティの観点から吟味する。

### 1) テレビ授業（インターネット配信TV授業）

総務省は視聴覚障害者や高齢者に配慮した字幕放送等の普及促進に努めており、2016年度NHKの総放送時間に占める字幕付番組は全体の84.4%である。放送大学のTV授業の字幕化も同様の方針が進められ、学内の放送部企画管理課素材管理係によって内製、外製を含め、手元のリモコンでON/OFFを操作できるクローズドキャプション型字幕が付与されている。2018年2学期には、TV授業科目の65.2%の字幕化が達成されており、今後、新規制作科目はすべて字幕化されるので、数年以内に字幕率100%に近づくと考えられている。インターネットで配信されているTV授業は

168科目である。TVで字幕化された科目がインターネット配信される場合にはすべて字幕が付与されている。

### 2) ラジオ授業（インターネット配信授業）

ラジオ授業の情報保障としては、聴覚障害等の学生から要望があれば、可能な限り講義台本を渡すことになっている。しかし、実際には完全な形で講義台本がある授業の方が少ない。この理由については3節において説明する。2018年2学期には178科目のラジオ授業がインターネットで配信されている。その中で、9科目のみが実験的に字幕・静止画が付与されたコンテンツとしてインターネット配信されている。

### 3) オンライン授業

2016年度に開講したオンライン授業は、ほとんどの科目に字幕が付与され、国内の大学、通信制大学のオンライン授業の中でも音声の情報保障としてはトップクラスのアクセシビリティの高さを誇っている。（音声認識技術の活用については、本稿の後半で紹介する。）

## 3. ラジオ授業のアクセシビリティの確保にむけて

### 1) ラジオ授業の情報保障と障害者支援プロジェクトの取り組み

前章でラジオ授業の台本が完全な形で残っている科目が少ないと記したが、その理由として、スタジオ収録時にしばしば台本の修正、カット、再構成が行われ、出来上がった番組が、当初の台本と一致しない場合が多いことが挙げられる。現状では完成された授業に則して台本を整える作業は、制作側にも講師側にも課されていない。学生課によれば、学生からの台本の要求は年間6件程度と少ないがすべての要求に応えられるわけではないという。PEPNet-JAPAN<sup>3</sup>の活動に参加している著者は、単位互換校である筑波技術大学や全国の大学等の聴覚障害支援の関係者に、本学のラジオ授業に関するインタビュー調査を試みているが、ほとんどの学生が印刷教材だけを頼りに学習していると答えており、ラジオ授業の情報保障が、障害者差別解消法の合理的配慮の観点から喫緊の課題であることは明らかである。

### 2) ラジオ授業のインターネット配信コンテンツへの字幕・静止画付与

本学の情報コースの障害者支援プロジェクトでは、2012年よりラジオ授業のアクセシビリティ向上のためにインターネット配信に着目し、以下のような試みを実験的に行ってきた。

現状のインターネット配信のラジオ授業は、TV授業と同じ様式の画面が表示されるが、画面は黒いまま音声のみが流れてくる仕組みである。そこで著者



ラジオ番組字幕付与実験画面

は、2013年度開設科目ラジオ特別講義「メディアと与謝野晶子」（講師 松村由利子・広瀬洋子）に、著作権フリーの写真や静止画を挿入した、講義音声と同期させて字幕を画面に付与するコンテンツを制作し、本学の学生向けインターネットサイトから実験的に配信を開始した。聴覚障害学生及び関係者からは、初めてラジオ授業を見ることができたと好評であり、放送大学の正規の配信コンテンツとして安定的に供給するために、コストの低減、講師やコーディネータの作業分担等の最適化などに関する研究開発をその後も続けてきている。その中で、後述する音声認識技術の活用も行っている。2018年度現在、学部4科目、大学院3科目、特別講義2科目、合計9科目のコンテンツが配信中である。この字幕・静止画付与の映像教材の研究開発については別の論考で発表する予定である。

字幕・静止画付与配信中のラジオ授業科目<sup>4</sup>

#### 教養学部

- (1) リスク社会のライフデザイン (14')
- (2) 心理臨床の基礎 (14')
- (3) CGと画像合成の基礎 (16')
- (4) 環境問題のとらえ方と解決方法 (17')

#### 大学院

- (1) データベースと情報管理 (16')
- (2) 音楽・情報・脳 (17')
- (3) 情報学の技術 (18')

#### 特別講義

- (1) メディアと与謝野晶子 (13')
- (2) 急増するうつ病 (17')

## 4. 放送大学のラジオ教材の字幕化への道のり

### 1) 音声認識技術を活用した京都大学との共同研究と実践<sup>5</sup>

著者は2011年から音声認識技術を用いて、放送大学の講義の字幕付与を効率的に行う方法について検討を

進めてきた<sup>6</sup>。

2012年から京都大学学術情報メディアセンターの河原達也研究室と共に、音声認識技術を用いて、放送大学の講義をテキスト化し、字幕付与を効率的に行う方法について研究を進めてきた。深層学習などで進化させてきた京大の最新技術と本学の高品質の音源とテキストデータが会うことによって、話者によっては90%程度の認識率を実現させた。その上で音声認識結果を編集することにより、字幕テキストを効率的に作成できるかを検証した。

音声コンテンツの字幕化は、従来、人の手によるテープ聞き起こし作業の延長線上にあるものであった。放送大学のTV授業の字幕化のコストは10年程前には、45分で17万円という時代があった。その後、入札が行われるようになり、4～5万円台を推移するようになった。近年、音声認識技術は深層学習の導入などで大幅に性能が向上し、スマートフォン等での入力では実用的な水準に達してきている。しかし、話し言葉の音声認識はまだ容易でなく、講演や講義の字幕付与のための研究開発が進められている。

### 2) 音声認識・字幕生成サーバ 京都大学での検証

京都大学では、講演や国会審議<sup>7</sup>を対象とした音声認識の研究を進めてきており、そうした音声コンテンツに対して音声認識と字幕付与を行うサーバ (<http://caption.ist.i.kyoto-u.ac.jp/>) を構築している。

利用者は、音声ファイルや映像ファイルを当該サーバにアップロードし、所定の手続きをすると、音声認識による書き起こしにタイムスタンプが付与されたファイル (SAMIやSRTなど複数のフォーマット) が生成される。これらは、一般的な再生ソフトで字幕ファイルとして利用可能である。

ただし音声認識には誤りが含まれる上に、話し言葉には言い淀みなども多いため、字幕として提示するには編集が欠かせない。また適当な位置での改行や句読点挿入も必要である。そのためのエディタも上記サイトで提供している。

活用を想定しているコンテンツは以下の3種類で、各々には音声認識のモデルが用意されている。

- ・講演：学会や講義など大教室で1人で行う学術講演
- ・スピーチ：一般的な話題に関してゆっくり話すもの
- ・討論：議会審議など公共の場での複数人による討論

音声認識システムはすべて、JuliusとDNN-HMM音響モデル (系列識別学習) 及び単語3-gram言語モデルで構築している。また、コンテンツに関連するテキスト (予稿やスライドなど) を同時にアップロードすることで、音声認識の言語モデルを適応することも可能である。

### 3) 放送大学講義の音声認識

京大の音声認識技術の研究にとっても本学の授業音源の活用は、以下の点で有益なものであった。

- ① スタジオ収録された録音条件の良い音源

- ② 多くの場合、講師が台本を読むストレートトークで、一般の講演や講義に比べるとはるかに発音が明瞭（一部には対談形式やインタビュー等が挿入されている場合もある。）
- ③ 言語的には、講義特有の専門用語が数多くあるのが問題となるが、印刷教材のテキストデータが入手可能である。

音声認識の音響モデルは、『日本語話し言葉コーパス』（CSJ）の学会講演約260時間を用いて学習したDNN-HMMである（前章の講演カテゴリに相当）。放送大学の講義音声約60時間を用いて、半教師つき追加学習も試みた。

言語モデルは、同じくCSJを用いて構築されているが、科目毎に印刷教材のテキストを追加混合することで適応を行う。これにより専門用語の追加も自動的に行われるが、正しい読みが付与されているか確認が必要である。次節に示すように、概ね90%の認識率が得られている。

4) 音声認識を用いた字幕テキスト作成実験

いくつかの講義を対象に、音声認識を行ったものを編集することで字幕テキストを作成する実験を行った。これは、最終的に放送大学で字幕として配信するものでなく、テキストとして誤りがなく、適度に同期されて改行がされているレベルのものである。ここで実験に用いた講義は、2015年度に開講された以下のラジオ授業番組である。

- (1) 心理臨床の基礎
- (2) リスク社会のライフデザイン
- (3) CGと画像処理の基礎

講義はいずれも各回45分で15回あるが、(2)については台本のある3回分を除いた。また、(3)については本実験では8回分のみを用いた。

平均の音声認識率を表1に示す。これは、最終的に生成された字幕テキストに対する文字単位の正解率である。この編集に要した時間と実時間比も表1に示す。これは、1名の作業者が前述したエディタを用いて行ったものである。この後、実時間と同程度の確認作業を行っている。また講義(1)(2)について、各回の認識率と編集時間の関係をプロットしたものを図2・3に示す。

図2・3から音声認識率と編集時間の間には、かなり高い(0.5~0.6)相関があることがわかる。これ以前に放送大学で行った実験では、音声認識を用いずに放送大学の講義を書き起こした場合、実時間の平均5.3倍(約4時間)程度と報告されている<sup>6</sup>。これをグラ

表1 講義の音声認識率と編集時間

	講義数	認識率	編集時間	実時間比
(1)	15	90.8%	3時間16分	4.4
(2)	12	88.5%	3時間46分	5.0
(3)	8	94.4%	3時間29分	4.6

フ中の直線と重ねると、87%程度の認識率の場合に相当する。これから、音声認識率が87%以上の場合にその効果があることが示唆される。また、93%になると1/3以上の時間短縮効果が示され、かなり優位性があるといえる。

前記の3つのラジオ科目は、2016年度から本学のインターネットサイトから字幕と静止画が付与されたコンテンツとして実験配信が行われている(図4参照)。

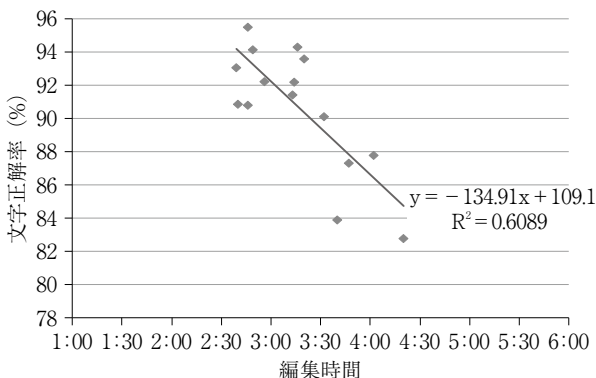


図2 音声認識率と編集時間の相関(「心理臨床の基礎」)

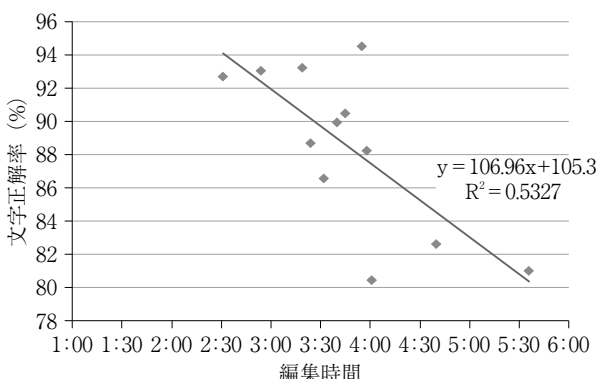


図3 音声認識率と編集時間の相関(「リスク社会のライフデザイン」)

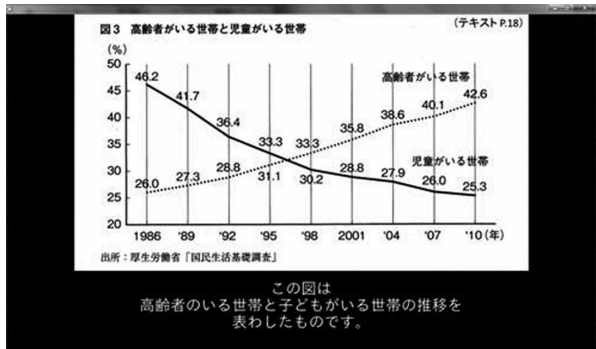


図4 ネット配信の(ラジオ)講義への字幕付与例(「リスク社会のライフデザイン」) ©放送大学

## 5. 放送大学のオンライン授業の字幕化

### 1) 音声認識を用いたオンライン授業の字幕付与実施事例

2016年度から開設されたオンライン授業では、前述した実験を踏まえ、当初から、音声認識エンジン Julius を用いて字幕が付与された (図5 参照)。エディタは「おこ助」字幕入力ソフトを活用した。Julius を活用して字幕付与が行われたオンライン授業は合計 115 時間、12 科目に及んだ。

[2016年度開講]

- (1) がんを知る (15回)
- (2) 女性のキャリアデザイン入門 (8回)
- (3) 感性工学入門 (8回)
- (4) メディアと知的財産 (15回)
- (5) 物理演習 (8回)
- (6) 臨床推論 (8回)

[2017年度開講]

- (7) 学校と社会を考える (15回)
- (8) データの科学 (15回)
- (9) フィールドワークと民族誌 (15回)
- (10) 生涯学習を考える (15回)
- (11) 女性のキャリアデザインの展開 (8回)
- (12) イランとアメリカ (15回)

オンライン授業は以下のサイトで公開されており、体験版を視聴することもできる。

<http://online-open.ouj.ac.jp/>

### 2) 字幕制作の課題：オンライン授業の制作現場から

ここで本学のオンライン授業の制作における音声認識技術を活用した字幕制作過程を整理し、本学特有の課題を抽出したい。

オンライン授業の字幕制作過程

- ① オンライン授業に用いる仮編集の動画を受け取る。



図5 オンライン授業への字幕付与例 (「物理演習」 ©放送大学)

- ② 京大の「字幕自動生成システム」(<http://caption.ist.i.kyoto-u.ac.jp/caption/>) にデータを送信し、src形式のテキストデータを得る。
- ③ オンライン授業上では、srcではなくvttを使用しているため、データ変換を「おこ助」字幕入力ソフトで行う。
- ④ テープ起こしの会社に動画と字幕vttデータを渡し、内容の修正を依頼する。
- ⑤ 納品データを「おこ助」で取り込み、字幕の長さの調整 (1文当たりの文字量の調整) や、タイミングの調整、内容の確認を行う。
- ⑥ 上記の字幕をオンライン授業にのせる。

Julius活用の利点としては、第一にフリーソフトでコストがかからないこと、第二に、音声タイムコード付きでテキストデータを入手することが出来ることである。音声認識率は、話者の話し方により高低があり、いずれにしても、人による確認と修正は不可欠である。

課題としては、第一にオンライン授業は、ラジオ授業のように印刷教材のテキストデータは存在しないので高い音声認識率を望めないこと、第二に、放送大学で専用のサーバを運用していないので、Web経由頼みとなっていたことである。たとえば、本学のネットワークの不調でデータが送れない場合や、京都大学のサーバ側に問題がある場合に備えて、常時安定的にデータが送付できる環境を整備する必要がある。音声認識技術一般に対して言えることであるが、講師の話し方や声のトーンなどによって、音声認識技術との相性があるので認識率にばらつきがある。コストの面から考えると、音声認識技術の活用に向けた講義と、最初から人手による字幕入力に適した講義を振り分ける必要もあると考えられる。

## 6. まとめ：本学の情報保障と今後の課題

本稿では、本学のメディア教材のアクセシビリティ向上のための字幕付与に関わる研究開発の推移にも触れてきたが、この間の音声認識技術の進歩は目覚ましいものがある。京都大学との共同研究によって開かれた本学の音声認識技術の活用は、進化する技術に導かれて次のステップに入りつつある。

その一つがUDトーク<sup>8</sup>との出会いである。本学では、面接授業として年間3,000以上の講座が全国各地の学習センターで開講されるが、聴覚障害者のために、学習センターによっては、事前に講師のレジュメやPPTを学生に渡すことはあったが、いわゆる合理的配慮としての情報保障はなされてこなかった。自分でノートテイクをつけて受講している者もいたが、面接授業も受けないケースの方が多かったと思われる。こうした状況を改善するために、本学は2016年にUDトークと法人契約を結び、以来、高知学習センター等の面接授業で聴覚障害のある学生の情報保障に活用されてい

る。UDトークは、音声認識エンジンAmiVoiceを搭載した市販の音声認識ベースの文字起こしソフトウェアであり、当初は聴覚障害者のその場の情報保障のために開発された。コストはかかるが法人契約の場合は、専用のサーバを建てるのが可能で、単語登録することで高い音声認識率を実現することができる。面接授業の場合は、準備段階で専門用語を登録し、講師が高性能のマイクを用いれば、学生は手元のiPad等の端末で教室内のどこにいても、間違いの少ないテキスト化された講師の講義を享受することができる。

オンライン準備室では、このUDトークが従来使用してきた字幕入力ソフト「おこ助」と連動していること、また目の前で文字起こしの状況を確認することができるという利点もあって、2016年から2017年にかけてJuliusの活用からUDトークの活用に移行し現在に至っている。UDトークの活用に関する研究開発に関しては次の機会に論考として発表する予定である。

字幕は聴覚障害者のための情報保障として必要なものであるが、聴覚に障害がない一般の学習者にとっても有益なものである。大学レベルのアカデミックな講義の内容を音声のみで理解し、記憶に定着させることは至難の業である。専門用語や固有名詞を字幕によって学習することで、一層の学習効果が見込める。またUDトークに搭載された翻訳機能を活用すれば多言語化も可能である。国の内外を見渡しても、本学ほど高品質なメディア教材を数多く持っている高等教育機関は他に類を見ないだろう。こうした教材を簡便に迅速に多言語に字幕化し、留学生を含む国内外の学生に提供することは、日本の高等教育にとっても、社会全体

にとっても国際化に向けて大きな推進力となるだろう。情報技術を活用した障害者支援を推進していくことは、少数の弱者救済のために留まらず、広い意味で新しい教育の手法を生み出し、教育の質の向上に貢献していく事に繋がっていることを、まとめの言葉としたい。

## 参考文献

- 1 障害を理由とする差別の解消の推進に関する法律施行令により国公立大学と同様に義務化が求められている。
- 2 広瀬洋子 (2011) 「放送大学における視覚障がい者への支援と課題」『放送大学研究年報 第29号』P.89-102
- 3 日本聴覚障害学生高等教育支援ネットワーク  
<http://www.pepnet-j.org/>
- 4 放送大学、システムわかば⇒放送授業のインターネット配信⇒MENU⇒ラジオ番組の字幕付与実験
- 5 河原達也、秋田祐哉、広瀬洋子 (2016) 「自動音声認識を用いた放送大学のオンライン授業に対する字幕付与」情報処理学会研究報告  
\* 第4章は本論文の一部を再構成したものを掲載している。
- 6 長妻令子、福田健太郎、柳沼良知、広瀬洋子 (2012) 「クラウドソーシングを活用した効率良い字幕作成手法」信学技報、WIT2012-25
- 7 秋田祐哉、三村正人、河原達也 (2015) 「音声認識を用いた講義・講演の字幕作成・編集システム」情報処理学会研究報告、SLP-108-2
- 8 「UDトーク：コミュニケーション支援・会話の見える化アプリ」 <http://udtalk.jp/>

(2018年11月5日受理)