

講演・講義の音声から字幕を付けるシステムを開発 —放送大学の講義で90%以上の認識率—

概要：放送大学では様々な講義を配信していますが、聴覚障害者等の情報保障のために字幕付与を進めています。京都大学では講演・講義を対象とした自動音声認識の研究開発を進めており、最新の深層学習を用いることで、放送大学の講義に対しても概ね90%の認識率を実現しました。これにより、人手で書き起こしを作成するよりも効率的に字幕付与できることを確認しました。本システムは、放送大学で今年度から開始されたオンライン授業の字幕作成に活用されています。これほど組織的にこのようなシステムが活用されているのは初の事例です。本研究成果は、12月2日に情報処理学会アクセシビリティ研究会(SIG-AAC)にて発表されました。

今年度から施行されている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ており、聴覚障害者に対しては手話や字幕付与などの情報保障を行うことがこれに該当します。近年、様々な講義コンテンツがインターネット配信されていますが、字幕が付与されているものはほとんどありません。

放送大学は、我が国で最大のメディアを利用した高等教育機関であり、約300の科目の講義がテレビやラジオで配信されています。その大半がインターネットでも配信され、スマートフォンやタブレットなどでも視聴できます。現在、字幕が付与されているのはテレビ講義番組の半数程度ですが、近い将来100%の字幕付与を目指しています。今年度より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」も開設されており、原則的に字幕を付与する方針です。障害者支援において先進的な米国においても、オンライン学習の字幕は充実しているとはいえない現在、これは画期的なことです。ただし、人的・金銭的なコストが課題となっています。

河原達也 京都大学大学院情報学研究科教授、秋田祐哉 経済学研究科講師らは、自然な話し言葉を対象とした音声認識の研究を行っており、講演・講義に字幕付与を行うシステムを開発しました。河原教授らの音声認識技術は、2011年度（平成23年度）から衆議院の会議録システムにも導入されています。聴覚障害者のための字幕付与技術に関して公開のシンポジウムを毎年開催し、障害者や速記者・要約筆記者などと意見交換、及びシステムの実証実験を行ってきました。

講演や講義への字幕付与の形態には、その場でリアルタイムに行う場合と、収録した映像に事後的に行う場合があります。前者はリアルタイム性が重視されるのに対して、後者では高い正確性が求められ、人手による修正を想定しても、かなり高い音声認識精度でないと実用性がありません。特に大学の講義では、通常の音声認識システムでカバーされていない専門用語が多いことも課題でした。

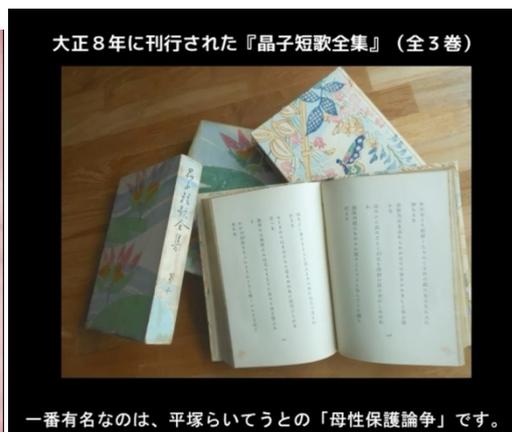
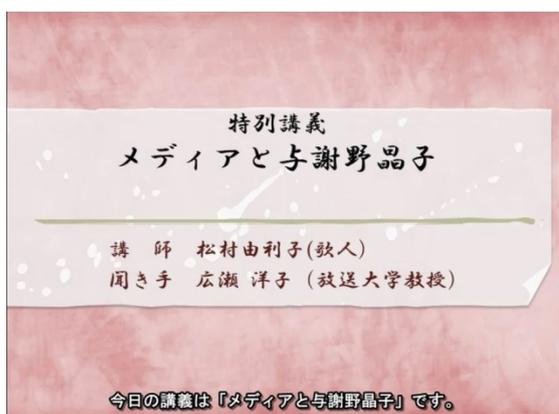
河原教授らは放送大学の広瀬洋子教授らと連携し、放送大学の講義を対象とした音声認

識・字幕付与に関して研究開発を進めてきました。大規模な講演・講義のデータベースを用いて最新の深層学習を導入し、さらに教科書テキストから専門用語などの表現を自動的に登録することで、概ね90%の認識率を実現しました。

約30の講義を対象に音声認識結果を編集する場合と人手ですべて書き起こす場合とを比較した結果、システムを用いることで作業時間が短くなることも確認しました。具体的には、システムの認識率が87%以上を超えると優位性が見られ、93%になると1/3以上の時間短縮効果が確認できました。

放送大学では、今年度から開始されたオンライン授業の字幕作成に活用されています。また、インターネット配信によるラジオ講義に字幕と静止画を付与したコンテンツも実験的に配信されています。

今後、他の教育機関で作成される様々な講義コンテンツに対する字幕付与にも展開されることが期待されます。



ラジオ講義のインターネット配信「特別講義 メディアと与謝野晶子」への字幕付与の例

<参考>

- ・ 情報処理学会 アクセシビリティ研究グループ(SIG-AAC) <http://ipsj-aac.org/>
- ・ 『聴覚障害者のための字幕付与技術』 シンポジウム <http://sap.ist.i.kyoto-u.ac.jp/jimaku/>