

高等教育における障害学生のための授業支援

～ 音声認識による字幕編集システムの概要～

Lecture Assisting Used by Speech Recognition for the University Student with Disability

長野大学 社会福祉学部

伊藤英一、石原剛志、旭 洋一郎

日本 IBM アクセシビリティ・センター

荒川健一、宮本 晃太郎、根岸則子

Keywords : 障害学生支援、音声認識、e-ラーニング

1 はじめに

障害のある学生が一般の学生と同じ授業を受けるためには、それぞれの障害に応じた授業支援が必要である。例えば、聴覚障害の場合には、要約筆記や手話通訳などのように、教員の話す内容（音声情報）を通訳者が文字や手話（視覚情報）に翻訳することで情報保障をおこなっている。

音声を機械的に文字化する自動翻訳技術としての音声認識は、同音異義語やかな漢字混じり文のような日本語の特性から文字化する際の誤変換が多い。そのため、訓練された別の話者が復唱して入力するシステム¹⁾や、複数端末から誤変換の修正をするシステム²⁾などある。このように日本語音声認識による即時変換には問題があるが、音声をテキスト（電子的な文字情報）に変換するため、その後の多様な利活用が期待できる。

そこで、認識率の向上や即時修正の技術ではなく、誤変換の多い音声認識の認識結果を容易に修正できるシステムを用いて、授業中のスライド画像や教員の音声などに字幕情報を付帯させることの利点に注目した。音声認識の認識結果であるテキストを有効に活用する一例として、授業支援（音声の字幕化）およびe-ラーニング（自習システム）という、より多くの学生が利用できるシステムの試験運用を開始したので概要を報告する。

2 システムの概要

長野大学と日本アイビーエムとの共同プロジェクト「障害のある学生のための授業支援プロジェクト」では、カナダの Saint Mary's University, Liberated Learning Project³⁾における音声認識を利用した授業支援システム ViaScribe(IBM)を参考に日本 IBM 東京基礎研究所アクセシビリティ・センターが開発した

字幕編集システムを利用して、聴覚障害学生が受講する授業においてシステムを運用し、その有効性及び運用方法などに関する問題点の明確化を目的としている。

また、一般的な聴覚障害学生の授業支援としてのノートテイク（要約筆記）に置き換えるものではなく、システムにより音声テキストに変換されることの有効性に着目し、効果的な活用事例を検討することで、より多くの学生を対象とした授業支援へと発展させることが大きな狙いである。

2.1 システム構成

システム構成図を図1に示す。まず、教員の音声をワイヤレスマイクにて受信し、教室の音響設備とノート PC のサウンド入力へ分配する。ノート PC では音声認識システム ViaVoice によりテキストに変換され、誤変換を含んだ状態ではあるが一旦聴覚障害学生へ提示される。また、同ノート PC は授業用スライド（パワーポイント）を提示するためにも利用する。

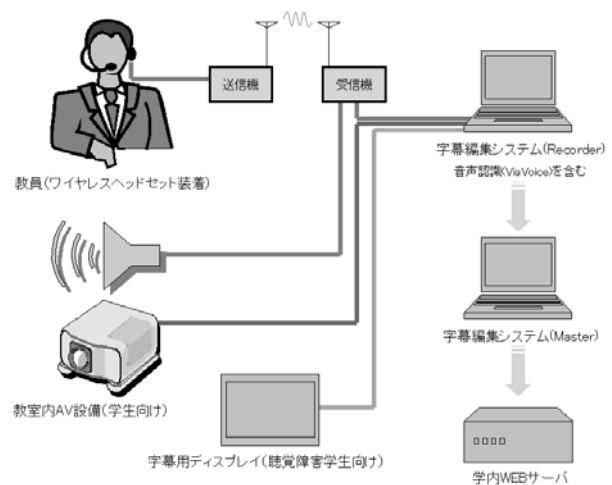


図1. システム構成図

字幕編集システム(Recorder)は、(1)教員の音声(wav)と、(2)音声認識結果(text)、(3)教室のスクリーンに表示されたスライド画像(gif)を授業90分の時系列情報として収録される。

その時系列情報は字幕編集システム(Master)により、授業終了後において音声とスライド画像などを参考に、誤変換の修正を行う。修正後は、音声とスライド画像に字幕を合成したマルチメディア情報(SMIL)に変換され、学内サーバへ格納される。

2.2 編集作業の流れ

時系列情報としての講義データ(教員の音声、音声認識結果、スライド画像)は、字幕編集システム(Master)で、編集者により適切な文字情報に修正される。編集画面を図2に示す。タイムスタンプの付いた未修正テキストが任意の文字数ごとに表示され、教員の音声やスライド画像をもとに文字列の確認をする。正しい、あるいは修正後は[]、表示しないは[x]、不確定は[]というように選択する。文字列の修正は、逐次でも事後でも可能である。

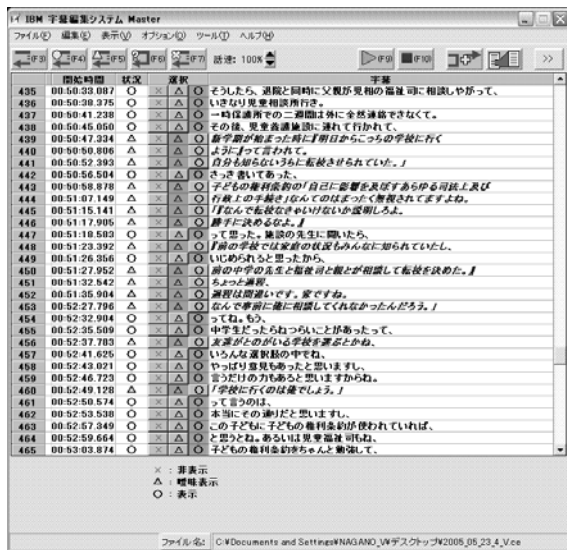


図2. 字幕編集システム(Master)の編集画面

編集後、テキストは授業中の教員の音声と、提示されたスライド画像と共に同期した時系列情報としてSMILファイルとして保存される。今回運用している授業は講義であることから動画像は取り込んでいないが、動きのある演習科目などでは動画像データも同時に提供することが可能となっている。

SMILファイルは、学内WEBサーバに格納してい

るため、学生は学内ネットワーク(無線LAN)にアクセスし、授業日ごとの講義データ(スライド画像と音声情報、字幕情報)を必要に応じて参照することが可能となる。(図3参照)

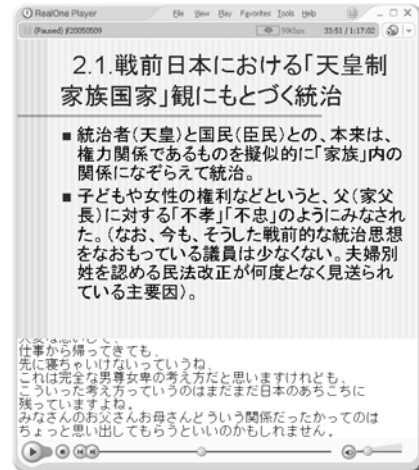


図3. 講義データの Window 画面

3 まとめ

4月より児童福祉論(登録学生308名)において本システムを運用し、聴覚に障害のある学生3名への授業支援と共に、筆記が困難な肢体不自由や視覚に障害のある学生、授業を欠席した学生に対するコンテンツとして配信している。

児童福祉論を開講している教室は、本学で最も大きく古い階段教室であるため、学生たちの私語はよく響き、かつ残響などの音響的問題も多いことから、教員が発声していない状態でも文字に変換される。さらに話し言葉であるため、声の強弱や強調、話速も不安定となり誤変換はさらに多くなる。しかし、ノートテイクがついている障害学生であっても、誤変換の多い無修正字幕も参照していることが散見されることから、ノートテイクと音声認識の用途の違いや、特徴についても調査してゆきたい。

参考文献

- 1) Liberated Learning Project:
<http://www.liberatedlearning.com/>
- 2) 音声同時字幕システム:
<http://www.bug.co.jp/products/onsei.html>
- 3) 遠隔地リアルタイム字幕提示システム
<http://www.tsukuba-tech.ac.jp/ce/fcs01/>

JOIN プロジェクト <http://www2.nagano.ac.jp/ito/join/>