

Raspberry Pi を活用した遠隔授業の音声認識システムの研究

Research on the speech recognition system for remote class using Raspberry Pi

古澤 太清¹⁾

指導教員：三輪 賢一郎¹⁾

1) サレジオ工業高等専門学校 機械電子工学科 情報コミュニケーション研究室

キーワード：音声認識, Julius, 発音辞書, 遠隔授業, Raspberry Pi

1. はじめに

昨今の新型コロナウイルス流行によって、大学等の教育機関において遠隔授業が本格化している。その遠隔授業をより効果的なものにするための一つの方策として、音声認識技術を用いて講義動画からテキスト情報を抽出したり、さらには字幕を付与したりすることが考えられる。

例を挙げると、TV放送等を通じて授業を実施している放送大学では、主に聴覚障害者への情報保障の一環として、音声認識技術を用いた講義音声のテキスト化・字幕付与に取り組んでいる[1]。また、留学生向けに、講義音声を音声認識した上で翻訳して英語字幕に変換するための研究も行われている[2]。しかしながらこれらのシステムは、ある程度のリソースや処理性能を備えた計算機を前提とした比較的大がかりなシステムであり、自宅での学習支援のツールとして全学生に貸与したりすることは難しい。

そこで、本研究においては、可搬性や経済性に優れた小型の計算機である Raspberry Pi (図1) に音声認識システムを実装し、Raspberry Pi 上で講義音声から音声認識によりテキストを抽出することを検討する。これにより、遠隔授業を受講する学生の学習支援に資するような、小型軽量で使い勝手のよい音声認識システムを構築することを狙う。



図1 本研究で使用した Raspberry Pi

2. システム概要

本研究で用いる音声認識エンジンとしては、限られたハードウェアリソースでも動作が可能で、かつフリーソフトである Julius を採用した[3]。ただし Julius には Raspberry Pi 用のバイナリファイルは用意されていないため、Raspbian OS 上に Julius のソースをコンパイルして実装した。また、音声認識に必要な音響モデル、発音辞書、及び言語モデルについては、Julius のディクテーションキットに同梱されているものを用いた。本研究で用いたソフトとハードの諸元を表1と表2に示す。

表1 本研究で使用した音声認識システム

項目	摘要
認識エンジン	Julius4.5
音響モデル	DNN-HMM 音響モデル
発音辞書	64,274 語
言語モデル	『現代日本語書き言葉均衡コーパス』ベース

表2 本研究で使用したハードウェア

項目	摘要
ハードウェア	Raspberry Pi 3B+
OS	Raspbian ver10.4

また、Julius で読み込み可能な音声ファイルのフォーマットを表3に示す。

表3 Julius の音声ファイルフォーマット

項目	摘要
ファイル形式	WAVE ファイル
サンプリング周波数	16 [kHz]
チャンネル	モノラル
上限ファイルサイズ	320,000 サンプル (20 秒)

3. 動作確認試験

前節の方法で構築したシステムの動作確認試験を実施した。事前に準備した短文の音声ファイルを Julius に入力し、音声認識結果（「少々お待ちください。」）が正しく出力されることを確認した。（図 2）

```
STAT: 00 _default: 12 sentences have been found
STAT: 00 _default: 4910 generated, 1480 pushed, 329 nodes pop
sentence1: 少々 お 待ち ください。
wseq1: <s> 少々+名詞 お+接頭辞 待ち+動詞 ください+動詞 </s>
phseq1: sp_S | sh_B o:_I sh_I o:_E | o_S | m_B a_I ch_I i_E
cmscore1: 0.495 0.988 0.969 0.966 0.692 1.000
score1: 271.036407
```

図 2 動作確認 出力結果

4. 実験計画

本節では、今後実施する実験の内容について示す。

今回の実験に用いる遠隔授業の講義素材は、2020 年前期に開講された「通信工学(3年)」と「電気通信法規(5年)」の2科目で、いずれも同一の男性教員によるものである。講義素材は MP4 フォーマットの動画ファイルであるので、それを所定の音声ファイルのフォーマットに変換した上で、構築済みの音声認識システムに読み込ませて音声認識を実行する。認識精度の評価指標としては、音声認識結果の文字認識率で評価を行う。文字認識率の算出式を次に示す。

文字認識率

$$= \frac{\text{正解文字数} - \text{誤挿入文字数} - \text{誤削除文字数} - \text{誤置換文字数}}{\text{正解文字数}}$$

具体的には、次に示す3段階にて実験を実施する予定である。

(1) ディクテーションキットに同梱されている発音辞書(64,274語)を用いて講義素材の音声認識を実行し、認識精度の評価を行う。

(2) 認識精度を向上させるために、講義内で出現すると考えられる専門用語を発音辞書に追加した上で講義素材の音声認識を実行し、認識精度の評価を行う。

(3) 提案法の絶対的な認識精度を確認するために、YouTubeの字幕機能を用いて同じ講義素材の音声認識を実行し、提案法との認識精度の比較評価を行う。

4. 結論

本研究においては、コロナ禍の状況で遠隔授業を受講する学生の学習支援に資するような、小型軽量で使い勝手のよい音声認識システムの構築を検討し、その基本的な動作確認までを実施した。引き続き、本システムの認識精度の評価実験を計画に沿って実施する予定である。

謝 辞

本研究には、汎用大語彙連続音声認識エンジン Julius[3]、ならびに国立国語研究所の『現代日本語書き言葉均衡コーパス』(BCCWJ)を利用した言語モデルを同梱した Julius ディクテーションキットを利用しています。

参考文献

- [1] 河原達也, 秋田祐哉, 広瀬洋子, "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, Vol.2016-AAC-2, No. 5, 2016年12月
- [2] 須藤克仁, 林輝明, 西村優汰, 中村哲, "授業アーカイブの翻訳字幕自動作成システムの試作," 情報処理学会研究報告, 2019-NL-240(15), pp.1--4, 2019年6月
- [3] 汎用大語彙連続音声認識エンジン Julius プロジェクト Web サイト (<https://julius.osdn.jp/>)