

サレジオ高専における学校案内システム向け 音声認識システムの研究

Research on speech recognition system for the campus information system
at Salesian Polytechnic

情報コミュニケーション研究室

江里口俊亮¹⁾

指導教員 三輪賢一郎¹⁾

1) サレジオ工業高等専門学校 情報コミュニケーション研究室

キーワード：Raspberry Pi 4, Julius, 音声認識, 学校案内

1. 背景

近年、音声アシスタントシステムや検索エンジンでの音声検索など、音声認識システムを活用した様々な機器やシステムが広く普及してきている。また、昨今の新型コロナウィルス流行により、直接的な接触を最低限に保ちつつ様々な用途に利用できる音声認識システムの需要は、今後さらに高まると考えられる。

音声認識システムの応用例の一つに、大学のキャンパス案内システムがある[1]。例えば、奈良先端科学技術大学院大学の音声対話ロボットは、音声による対話及び頭部や腕などのジェスチャによるインターフェースを用いて受付案内が行え、また眼球部分のカメラによる画像処理が可能なシステムである[2]。しかし、これらの機能を最大限に発揮するためには、かなり高性能なハードウェアに加え、ネットワークへの常時接続環境が必要となる。

本研究では、低価格かつ低消費電力で、必要最低限の性能で十分に動作する、新入生や外部からの訪問者に有用な、学校案内システム向けの音声認識システムを検討する。具体的には、小型コンピュータである Raspberry Pi に、オープンソースで配信されている音声認識エンジン Julius をインストールしてシステムを構築する。本システムは、音声認識を用いて本校（サレジオ高専）内の施設の位置や概要を簡易に検索できるシステムへの搭載を想定したものであるが、小型軽量かつ低

消費電力という特長を持つことから、学校案内という用途以外への応用も期待できる。

2. システム概要

2.1 Raspberry Pi

図1は本研究でハードウェアとして用いる Raspberry Pi 4 Model B の本体画像である。Raspberry Pi は、小型で安価かつ低消費電力で、拡張性にも優れていることから、本研究で想定するシステムの要求条件を満たすものである。



図1 Raspberry Pi 4 model B

2.2 汎用大語彙連続音声認識エンジン Julius

音声認識エンジンの Julius は大語彙連続音声認識を主な目的として開発されたフリーソフトウェアである[3]。本システムにインストールしたソースコードのバージョンは 4.4.2.1 である。Julius には混合ガウスモデル (GMM) とディープニュートラルネットワーク (DNN) の二種類の音響モデルが存在する。

また、GMM は認識速度が速いが認識精度は低く、DNN は認識速度が遅いが認識精度が高い。

表1 音響モデル性能

	認識速度	認識精度
GMM	速い	低い
DNN	遅い	高い

2.3 発音辞書

発音辞書は、音声を認識して文字として書き出すために必要な日本語の単語などのデータが収められている辞書ファイルである。発音辞書を含むディクテーションキットのバージョンは4.4で、辞書には約60,000語の登録データが存在する。

3. 実験計画

今後の実験計画を下記に示す。なお、予備実験を通してシステムが正常に動作することは確認済みである。

(1) 辞書登録

ディクテーションキットに含まれる既存の辞書（デフォルト辞書）に、サレジオ高専のホームページやパンフレットなどから抽出した70単語を追加する。

追加した単語の例)

- ・サレジアンホール
- ・夢工房
- ・情報館

(2) 音声録音

(1)で追加した70単語を、自分自身と協力者の計5人で発声し、録音する。

(3) サンプリング

取得した録音データのフォーマットを、Juliusで認識可能な16kHz・モノラルのWAVEファイルに変換する。

(4) 認識実験

表2に認識実験の組み合わせを示す。

①GMM&デフォルト辞書の組み合わせ

実験(3)でサンプリングした音声データを、GMMとデフォルト辞書（単語追加なし）の組み合わせを用いて音声認識を実行する。

②GMM&単語追加済辞書の組み合わせ

実験①の単語辞書を70単語追加済み状態に変更して音声認識を実行する。

③DNN&デフォルト辞書の組み合わせ

実験(3)でサンプリングした音声データを、DNNとデフォルト状態の単語辞書の組み合わせを用いて音声認識を実行する。

④DNN&単語追加済辞書の組み合わせ

実験③の単語辞書を70単語追加済み状態に変更して音声認識を実行する。

(5) 認識率計算

(4)の実験の①～④の認識結果からそれぞれの認識精度を算出する。

表2 認識実験の組み合わせ

実験番号	組み合わせ	
①	GMM	デフォルト 辞書
②	GMM	70 単語 追加済辞書
③	DNN	デフォルト 辞書
④	DNN	70 単語 追加済辞書

4. 結論

本研究では、低価格かつ低消費電力で、必要最低限の性能で十分に動作する、学校案内システム向け音声認識システムを検討した。今後は、実験計画に沿って本システムの評価を行い、システムの有効性を確認する予定である。

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た。

参考文献

- [1] 大浦圭一郎, 山本大介, 内匠逸, 李晃伸, 徳田恵一, “キャンパスの公共空間におけるユーザ参加型双方向音声案内デジタルサイネージシステム”, 人工知能学会誌, pp. 61–62, 2013年1月
- [2] 西村竜一, 内田賢志, 李晃伸, 猿渡洋, 鹿野清宏, ”Juliusを用いた学校案内ロボット用音声対話システムの作成,” 電子情報通信学会技術研究報告, 言語理解とコミュニケーション 101(520), pp. 93–98, 2001年12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/>)