長野県工技センター研報
No.17, p.E1-E4 (2022)

多品種生産に対応する音声認識協働ロボットシステムの開発*

油科　賢*1　坂本潤嗣*1　北村泰地*1

Development of Voice Recognition Collaborative Robot Systems for Multi-Variety Production

Satoshi YUSHINA, Junji SAKAMOTO and Taichi KITAMURA

音声認識ツールとしてJuliusとSpeechRecognitionを使用し，手先位置の移動や姿勢の記憶，姿勢の呼び出しなど基本的な動作を音声で指示することがきる協働ロボットシステムを構築した。

このようなロボットは工程の変更に伴い必要となるロボットの動作変更を容易にし，また逐一動作指示が必要になるようなロボットとの協働作業も効率化することができるため，多品種生産における工程の省力化に貢献できる。

キーワード：音声認識，Julius，SpeechRecognition，協働ロボット，多品種生産，省力化

1　緒　　言

県内中小製造業では以前より人手不足や生産性向上が課題となっているが，特に人手不足については今後の生産年齢人口の減少に伴いさらに深刻化すると思われる。これらの課題の対策の一つとして，協働ロボットの導入が進みつつある。協働ロボットはこれまでの産業用ロボットとは異なり安全柵が不要で人と同じ空間で作業ができ，人の代わりに製造ラインに入ることができるようなロボットである。また小型，省スペースであることからフレキシブルな生産ラインにも対応可能である。

一方，県内製造業に対しては多品種少量生産の要求が増加している。多品種生産では生産ラインの変更が頻繁に発生するため，それに応じたロボットの動作変更も必要になる。多品種生産の増加に伴いこのようなロボット動作の教示作業(ティーチング)も増加することからこれらの作業に対する省力化も検討していく必要がある。

人と協働ロボットの理想的な関係は，作業者が人と作業をするのと同じようにロボットに指示できることであると考える。それが可能になればティーチングなどの動作変更も容易になり，また状況に応じてその都度動作指示が必要になるような協働作業も効率的に行うことができる。そのためにはロボットが人の音声を認識し動作する必要があるが，現状製造現場でそのように使用できる設備は少ない。

人と同じように指示できる協働ロボットシステムの構築を最終的な目標とし，本研究では第一段階として音声で基本的な動作を指示することができる協働ロボットシステムを構築した。

2　システムの構成

2．1　協働ロボット

音声で操作するロボットには(株)デンソーウェーブ製の協働ロボット「COBOTTA」を使用した。図１に協働ロボットを，表１に協働ロボットの仕様を示す。

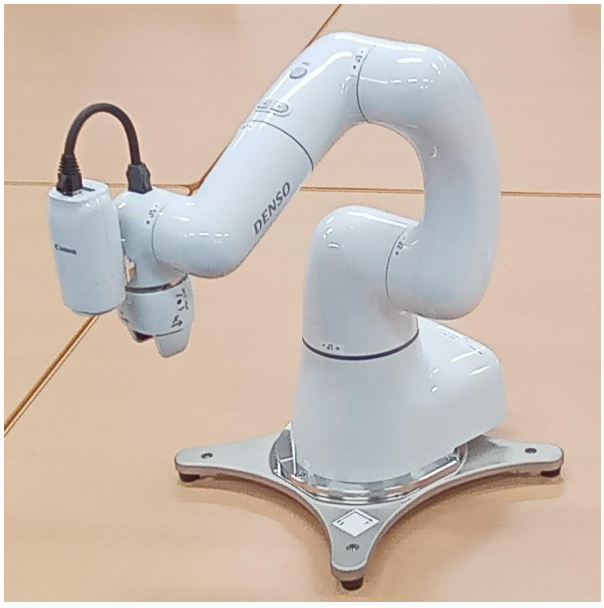

図１　協働ロボット
（デンソーウェーブ製 COBOTTA）

*　特別研究
*1　情報システム部

表１　主な仕様

軸数	6 軸(アーム部) ＋ 1 軸(電動グリッパ部)
アーム長 (第 1+第 2 アーム)	342.5(165＋177.5)mm
定格可搬質量	0.5kg
位置繰返し精度	±0.05mm
外部通信	Ethernet×1回線 USB×2回線 VGA出力×1ch
本体質量	約4kg

表２　対応エンジン／API

エンジン/API	日本語対応
CMU Sphinx	×
Google Speech Recognition	○
Google Cloud Speech API	○
Wit.ai	○
MicrosoftBing Voice Recognition	○
Houndify API	×
IBM Speech to Text	○
Snowboy Hotword Detection	×

ＣＯＢＯＴＴＡは一般的な協働ロボットと比較してアーム長が短く可搬重量も小さいため実際の生産現場での利用は限定的になると思われるが，小型でコントローラも内蔵しているため取り回しがよく，外部機器との連携も容易であるため，本研究に適した協働ロボットである。

２．２　音声認識ツール

音声認識技術はスマートフォンやスマートスピーカの入力機能として一般的に利用されており様々なものが存在するが，本研究ではオープンソースの音声認識ツールJulius[1),2)]とpythonの音声認識ライブラリSpeech Recognition[3)]を使用した。

いずれのツールも音声ファイル(wav形式)やマイクから入力された音声データを言葉として認識し，その結果を文字列(テキスト)として出力することができる。

２．２．１　Ｊｕｌｉｕｓ

Juliusは京都大学や名古屋工業大学などが研究・開発を行っている汎用大語彙連続音声認識エンジンである。オープンライセンスであり商用利用も可能である。

Juliusの最大の特徴は，単語辞書や言語モデル・音響モデルなどの音声認識の各モジュールを組み替えることで，小語彙の音声対話システムからディクテーションまで様々な幅広い用途に応用できることである。

オフラインでも利用できるためネットワークに接続できない環境にも対応できる。

２．２．２　ＳｐｅｅｃｈＲｅｃｏｇｎｉｔｉｏｎ

SpeechRecognitionはpythonの音声認識ライブラリとして公開されている。SpeechRecognitionライブラリ自体が音声を認識するのではなく，外部の音声認識エンジンやAPIを利用して音声認識を実現する環境を提供している。

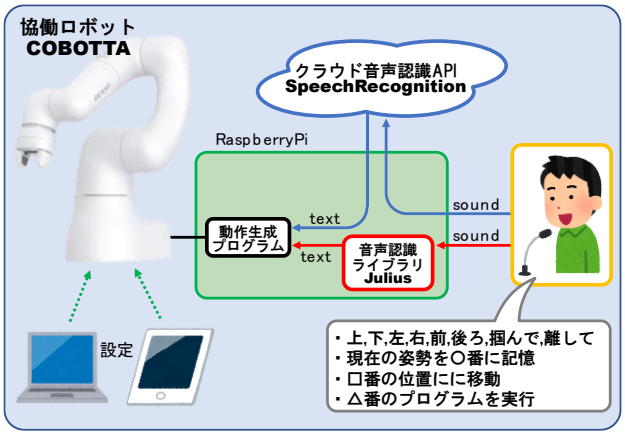

図２　システム概要

対応しているエンジン，APIを表２に示す。今回は音声認識エンジンとしてGoogle Speech Recognitionを使用した。

３　音声認識協働ロボット

３．１　システム概要

図２に構築した音声認識協働ロボットシステムの概要を示す。スタンドアロンで利用できるJuliusで音声認識を行う場合と，クラウド(Google Speech Recognition)で認識させる場合それぞれについてシステムを構築した。

作業者が発した音声はマイクにより入力され，クラウド音声認識を利用する場合はネットワークを介して入力され結果のテキストを受け取る。オフラインで利用する場合はローカルで処理が行われテキストが出力される。

その後認識された音声のテキストをプログラムによりロボットの動作に変換し協働ロボットを操作する。

一連の処理には小型ボードPCのRaspberry Pi4を使用した。

３．２　指示動作

音声認識協働ロボットの最終的な目標は自然言語から適切な動作を生成し，人が人に指示するように指示できることであるが，本研究ではその第一段階として以下の基本的な動作に対して音声操作を行うこととした。

(1)指示した方向へ一定量ハンド移動
(2)指示した方向へのハンド微小移動
(3)指定した方向へ距離を指定しハンド移動
(4)ハンドの回転
(5)現在の位置・姿勢を登録
(6)登録した位置・姿勢への移動
(7)保存された動作プログラムの実行

３．３　動作生成プログラム

音声認識ツールを介して得られたテキストは動作生成プログラムによりロボットの命令に変換される。このプログラムはテキストの中にあらかじめ決めておいたキーワード(方向，移動距離，動作の種類等)があるかどうか判定し，どのようなキーワードがどれだけ見つかったかに応じてロボット動作命令を出力するプログラムである。

４　結　　果

４．１　Ｊｕｌｉｕｓの認識精度

Juliusで音声認識を行うには認識のための音響モデルや言語モデルをパッケージ化した音声認識パッケージが必要となる。用途に応じた複数のパッケージが用意されているが，一般的な日本語の音声認識パッケージであるディクテーションキット(dictation-kit)を使用し認識精度を確認したところ，文章によっては頻繁に誤認識が発生した。例えば「おはようございます」という発話に対しては，10回中1～2回程度の誤認識があった。また「右に移動」という発話に対しては10回中5～6回程度の誤認識があった。認識結果の例を図３に示す。

この誤認識発生の原因の一つにディクテーションキットの辞書の精度がある。ディクテーションキットの辞書は，日本語を認識するために必要な情報からなり，この辞書のデータを充実させることで認識精度の向上が期待できる。しかしながらあらゆる日本語に対応できる辞書を個人的に作成するのは現実的に不可能である。よって今回は使用する言葉を限定することで認識精度を向上させる手法をとった。

Juliusには独自辞書を作成するツールが含まれており，これを使用することで決められた言葉に対してのみ認識を行うような辞書が作成できる。これにより作成した辞書ファイルを使用した場合「右に移動」という発話に対しては10回中10回とも正しく認識することができ，認識精度の向上が確認できた。表３に独自辞書の登録単語を，表４に登録した単語を組み合わせて作られる動作指示の文章例を示す。

ただしこの場合「おはようございます」という言葉は辞書に登録されていないため必ず誤認識となる。また，

```
pass1_best:  おはよう ござい ます 。
pass1_best_wordseq: <s> おはよう+感動詞 ござい+動詞 ます+助動詞 </s>
pass1_best_phonemeseq: silB | o h a y o: | g o z a i | m a s u | silE
pass1_best_score: -4575.377441
### Recognition: 2nd pass (RL heuristic best-first)
STAT: 00 _default: 9426 generated, 1547 pushed, 220 nodes popped in 178
sentence1:  おはよう ござい ます 。
wseq1: <s> おはよう+感動詞 ござい+動詞 ます+助動詞 </s>
phseq1: silB | o h a y o: | g o z a i | m a s u | silE
cmscore1: 0.417 0.794 0.922 0.968 1.000
score1: -4611.598145
```

(a)正しく認識

```
pass1_best:  こう いう こと あり ます 。
pass1_best_wordseq: <s> こう+副詞 いう+動詞 こと+名詞 あり+動詞 ます+助動詞 </s>
pass1_best_phonemeseq: silB | k o: | i u | k o t o | a r i | m a s u | silE
pass1_best_score: -3749.394775
### Recognition: 2nd pass (RL heuristic best-first)
STAT: 00 _default: 43793 generated, 2944 pushed, 416 nodes popped in 140
sentence1:  と いう ほど あり ます 。
wseq1: <s> と+助詞 いう+動詞 ほど+助詞 あり+動詞 ます+助動詞 </s>
phseq1: silB | t o | i u | h o d o | a r i | m a s u | silE
cmscore1: 0.122 0.669 0.009 0.197 0.582 0.739 1.000
score1: -3789.038818
```

(b)「おはようございます」を誤認識

```
pass1_best:  右 に 行く 。
pass1_best_wordseq: <s> 右+名詞 に+助詞 行く+動詞 </s>
pass1_best_phonemeseq: silB | m i g i | n i | i k u | silE
pass1_best_score: -3602.104004
### Recognition: 2nd pass (RL heuristic best-first)
STAT: 00 _default: 63510 generated, 3900 pushed, 636 nodes popped in 140
sentence1:  右 に 行こう 。
wseq1: <s> 右+名詞 に+助詞 行こう+動詞 </s>
phseq1: silB | m i g i | n i | i k o: | silE
cmscore1: 0.348 0.402 0.550 0.004 1.000
score1: -3646.288330
```

(c)「右に移動」を誤認識

図３　Julius による認識結果例

「右に移動」という文章に対して「右方向に移動」や「右に動いて」という表現も正しく認識できない。これは登録されていない文章が入力されたときには登録されている中で一番近いと思われる文章を候補とするため，たまたま正しく「右に移動」と認識する場合もあれば，誤認識となる場合もあり精度は低い。

このため，独自の辞書ファイルを作成する場合は１つの動作に対して想定される文章表現を網羅的に辞書に登録するか，あるいは１つの動作に対してただ１つの文章表現を当てはめる必要がある。前者は辞書ファイルの作成が煩雑となり，後者は動作に対するただ一つの命令を作業者が記憶する必要があり，人に指示するような感覚で指示することができなくなる。

あらゆる表現を想定して辞書を作成するのは難しいため，一つの動作に対する命令表現をある程度限定して辞書ファイルを作成する必要があることが分かった。

４．２　ＳｐｅｅｃｈＲｅｃｏｇｎｉｔｉｏｎの認識精度

音声認識エンジンにGoogle Speech Recognitionを使用し精度を確認したところ，会話として日常的に使用するような発話でもほぼ正確に認識することができた。クラウド経由で認識するため，Juliusと比較して若干の遅延があるようにも感じるが，ロボットの動作スピードと比較すれば問題となるほどではない。ただ，ネットワーク環境に起因すると思われる数秒の遅延がまれに発生することがあった。

表３　独自辞書登録単語

種別	単語
方向	上，下，左，右，前，後ろ
動作	上げて，下げて，移動，傾けて，倒して，回転，掴んで，離して，登録，プログラムを実行
距離	1cm，2cm，3cm，4cm，5cm
番号	1番，2番，3番，4番，5番
助詞	に，を，の

表４　指示文章とロボット動作例

文章	ロボット動作
「右」 「右に移動」	ハンドを右に5cm移動
「ちょっと右」 「右にちょっと移動」	ハンドを右に5mm移動
「右に3cm移動」 「3cm右」	ハンドを右に3cm移動
「右に回転」	ハンドを右に90度回転
「右に傾けて」 「右に倒して」	ハンドを右に30度傾ける
「3番に登録」	現在の姿勢を3番として記憶する
「3番に移動」	3番に記憶してある姿勢に移動する
「2番のプログラムを実行」	2番に記憶してあるプログラムを実行する

５ 考　察

JuliusとSpeechRecognitionではSpeechRecognitionの方が精度良く音声を認識することが分かった。ただし，Juliusでも，使用する言葉を限定することで認識精度を上げることが可能である。また今回音声認識パッケージとして一般的なものを使用したが，設定やチューニングを変えることで精度の向上も期待できる。Juliusの特徴は，言語モデルや音響モデルの各モジュールをタスクに合わせて設計・構築することが容易で，認識の目的に合わせた調整が可能なことである。そのため生産現場のようなノイズの多い環境などでも，環境にあったセッティングを行うことでうまく認識できる可能性がある。ただしそのためには，音響や言語モデルの専門的知識も必要となる。

SpeechRecognitionは設定不要で高い認識率を得ることができるがその反面，ユーザ側で調整できることがほとんどない。ネットワーク接続も必須であるため，使用できる環境は限定的になる可能性がある。

６ 結　論

人と同じように指示できる協働ロボットシステムの構築を目指し，第一段階として音声で動作指示することができるロボットシステムを構築した。その結果は以下のとおりである。

(1) 音声認識ツールとしてJuliusとSpeechRecognitionを使用し基本的な動作(ハンドの移動・回転・把持動作・姿勢の記憶・姿勢の呼び出し・プログラムの実行)を音声で指示できるシステムを構築した。

(2) 音声認識ツールは使用する環境(環境音やネットワークの有無など)に応じて選択することで様々な場面で適用できる可能性がある。

(3) 本システムを使用することで，ハンドの位置やアーム姿勢の構築が容易になり，ティーチング作業の省力化が期待できる。

今後は指示できる動作をさらに追加し汎用性を向上させるとともに実際の運用に基づいた動作試験を行い，操作性を高めていく。

参考文献

1） A. Lee, T. Kawahara and K. Shikano.. "Julius --- An Open Source Real-Time Large Vocabulary Recognition Engine". EUROSPEECH, p.1691-1694（2001）

2） A. Lee and T. Kawahara.. "Recent Development of Open-Source Speech Recognition Engine Julius" APSIPA ASC 2009, p.131-137（2009）

3） Anthony Zhang. "SpeechRecognition 3.8.1". PyPA. 2017-12-05　https://pypi.org/project/SpeechRecognition/,（参照日2022-05-13）.