

# 音声認識による接客ロボットの開発

声認識について今から解説する. 音声認識とはコンピューターを用いて音声を入力し入力した音声を文字にして示すことである.

## 目次

1	はじめに	1
1.1	研究の背景	1
1.2	研究の目的	2
2	飲食店での接客について	2
2.1	接客の仕事内容	2
2.2	研究として検討する仕事	3
2.3	今年度のロボット	3
3	Julius	3
4	Open-JTalk	4
5	TCP/IP	4
6	研究手順	4
7	研究結果	4
8	考察	5
9	おわりに	6
付録 A	今回使用したもの	6

## 1. はじめに

### 1.1. 研究の背景

今回の音声認識は食堂にて注文を取るロボットがある場合, 人件費の経費削減ができるため制作してほしいという要望があった. そこで音声でのやりとりをするためのシステムを究明する必要があると考え, 本研究を開始した. 食堂で注文を取るロボットに使用した音声を認識させるシステムについて今から述べていく.

### 1.2. 研究の目的

食堂にて PC を実装したロボットに音声認識をするため Julius, Open-JTalk, 認識すべき単語, マイクを準備し実際に店で禁煙か喫煙かの判断や「あちらの席へどうぞ」「こちらの席へどうぞ」のようなレベルの案

内をすることを目的とする。図 1 に音声認識をしようとしたときのイメージ図を示す。



Fig. 1 様子

## 2. 飲食店での接客について

### 2.1. 接客の仕事内容

お客さまから注文を取り、出来上がった料理や飲み物を運び、空いた皿の片付け、レジ業務を行うのが主な仕事である。本研究のために協力していただいた店でははじめに「いらっしゃいませ」と客をお出迎えする。「何名様ですか」と人数の確認をし、このときに禁煙か喫煙かの確認も行う。5人以上の客でなおかつ禁煙の場合は「ただいま席をお作りしますので少しお待ちください」と言う。水とメニューを置く。メニューを置くときに水が出せない場合はメニューのみ出す。黒板メニューと呼ばれるものがある場合はお知らせする。小学生の低学年以下の子供がいる場合はキッズプレートを勧める。また、水を出すときは「いらっしゃいませ」と言いながらお客様の前に置き小さな子供がいるときはグラスが手の届かないように親の前などに置く。客の荷物などでテーブルが塞がっている場合は端にまとめて置く。オーダーを取るとき「お決まりですか？」と声を掛ける。決まっていない場合でも早めに聞くことで聞き忘れやオーダーの機器遅れによる客の不満も未然に防ぐ事ができるので必ず行う。注文するものが決まっている場合は客のしているページをよく見て確認しながら伝票へ記入する。聞き取りにくかった場合や客が勘違いしていそうな場合は「こちらの～でよろしいですか？」や「こちらの～ですが」などメニューを手差ししながら説明や確認をしてミスを防ぐ。注文するものが決まってない場合は「ではお決まりになったらお呼びください。」と言う。そしてドリンクバーやパン、ライスの確認を行う。オーダーの確認を行うときは「オムライスセットがおひとつ、ステーキライスがおひとつでよろしいですか？」のように復唱して確認を行う。このとき、客の顔をみて確認を行う。ドリンクバー付の場合は「あちらからご自由にお取り下さい」と場所を説明しながら確認する。厨房にオーダーを出す場合は大きな声で「お願いします。ランチ・ワン。パスタセットスリーです。」と言う。伝票下一枚を伝票版につけて厨房に渡す。残り一枚はホールコーナーに貼る。シルバーセットを使用する場合は「失礼します」と言い、昼の場合はBセット以外のシルバーをかごに入れて客の前に持って行き、夜の場合は人数分のかごにおしぼりを入れて客の前に持って行く。料理を出すときは「お待たせしました、オムライスです。」と言いながら出す。出す場所がわからない場合は「お待たせしました、オムライスのお客様は？」と確認をし、置く。料理出しの帰りは他のテーブルに目を配る。食事の終わってる所から「お下げしてよろしいですか？」と声掛けして食器を下げる。伝票チェック時に料理が出る都度ホールコーナーにある伝票にチェックを入れる。出揃ったら客のテーブルの下の伝票を伏せて置いてくる。デザートがある場合は下の伝票をデザートコーナーに持って行き上の伝票はホールコーナーの下段に下ろす。デザートがある場合は食後の進み具

合に気を配り食事が終わったらデザートコーナーに「1 番にアフターお願いします」とオーダーする。客に呼ばれた場合、席まで早足で行き「はい、呼びですか。」とお伺いする。追加オーダーの場合オーダーを下の伝票に記入・復唱確認後、食べ終わっている客の食器を下げてくる。食後にデザートが付くメニューの場合、空いた食器をさげる時に遅くなったときの不安を防ぐために「デザートをお持ちします。」と声を掛ける。レジではまず「ありがとうございます」といい、「オムライスおひとつで 80 円です。」「1000 円お預かりします。」「140 円のお返しです。」の流れでお金のやりとりを行う。また、複数の場合は「日替わりがおひとつとオムライスがおひとつで 1720 円です。」「5000 円お預かりします。」「3000 円と 280 円のお返しです。」「ありがとうございました。」と言う。レジで会計をする際に別払いと思われる客の場合は「ご一緒によろしいですか？」と確認する。カットはお客様がいる場合お冷サービスをしながら「お下げします。」と声を掛けて下げる。帰られた後は重ねて下げる。

## 2.2. 研究として検討する仕事

前述では接客の仕事内容を挙げたが今回、接客をするロボットを制作するにあたり、店の入口付近で客がタバコを吸うかどうかの判断や「あちらの席へどうぞ」「こちらの席へどうぞ」の案内をすることが研究として検討する仕事となる。

## 2.3. 今年度のロボット

今年度のロボットは図 2 に示すロボットを使用して音声認識を行う。



Fig. 2 使用したロボット

## 3. Julius

Julius は、音声認識システムの開発・研究のためのオープンソースの高性能な汎用大語彙連続音声認識エンジンである。数万語彙の連続音声認識を一般の PC 上でほぼ実時間で実行できる。また、高い汎用性を持ち、発音辞書や言語モデル・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、様々な幅広い用途に応用できる。[1] Julius はコマンド 1 つで起動させることができ、オプションでサーバモードや予め録音しておいた音声を認識させるなど元から用意したプログラムの改良せずに済むため今回は音声をマイクから認識させ、オプションでテキストファイルに変換するために使用する。

## 4. Open-JTalk

Open-JTalk は、日本語テキストを音声に変換するシステムである。ここでは c 言語を用いて aplay コマンドを system 関数により Julius より得たテキストファイルの内容を音声として発するために使用する。

## 5. TCP/IP

今回、Julius を使用するにあたり、サーバーモードとして使用することで PC1 台の中のみで操作を行う事ができ、サーバとクライアントを TCP/IP 経由で繋ぐために C 言語を用いてコンパイルし、繋ぐことができるように改良した。

## 6. 研究手順

- (1) 食堂にて入口前に PC を実装したロボットを設置する。
- (2) 客が入ったことを認識したら「いらっしゃいませ」と音声を発する。
- (3) 「いらっしゃいませ」と発した直後に「おタバコお吸いになりますか?」と発する。
- (4) 「禁煙です」「吸わないです」「吸いません」と答えたとき、「あちらの席になります」と答える。
- (5) 「喫煙です」「吸います」と答えたとき、「こちらの席になります」と答える。

## 7. 研究結果

Julius で音声を入力し、認識することを確認した。そのときの様子を図 3 に示す。また、Julius をサーバモードとして動かした。この時の様子を図 5 に示す。サーバモードで動かす際に研究用に c 言語で作成したクライアントを julius とつなぎ、音声を認識させた。しかし文字化けが起こった。そのときの様子を図 6 に示す。Open-Jtalk では研究用として c 言語でプログラムを組み、必要な音声を出力することに成功した。そのときの様子を図 4 に示す。

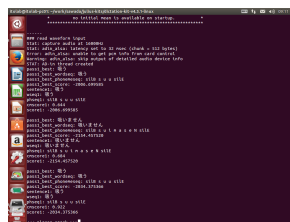


Fig. 3 認識結果

## 8. 考察

今回は Julius を使用して音声の入力や Open-JTalk を使用して音声の出力を行うことができたがそれぞれ独立して操作を行ったため、Julius を使用する際にテキストファイルを出力させ、そのテキストファイルを

Open-JTalk が音声として出力させるためのツールを見つけて使用することができれば実際に食堂にて使用することができると考えられる.

## 9. おわりに

今回ロボットに音声認識をさせることで人件費の削減につなげられるのではないかと考え, Julius や Open-JTalk で音声の入出力を行うことと決め Julius での音声の認識や Open-JTalk での音声の出力に成功した. 今後は Julius と Open-JTalk を連携する手段を見つけ, 人間とロボットの対話を行えるようにする.

## 参考文献

- [1] 第 7 章 言語モデル・[https://julius.osdn.jp/juliusbook/ja/desc\\_m.html](https://julius.osdn.jp/juliusbook/ja/desc_m.html)・2016/09/02 閲覧
- [2] Raspberry Pi で Julius を使った音声認識 (外部連携編)・<http://hyottokoaloha.hatenablog.com/entry/2015/07/03/131305>・2016/09/02 閲覧
- [3] mkdfa.pl リファレンス・マニュアル・[https://julius.osdn.jp/juliusbook/ja/mkdfa\\_pl.html](https://julius.osdn.jp/juliusbook/ja/mkdfa_pl.html)・2016/09/09 閲覧

## 付録 A. 今回使用したもの

yomi ファイルとは認識させるために準備した単語を書き込んだものである。今回使用した yomi ファイルを図 7 に示す。voca ファイルとはカテゴリごとに単語の表記と読み（音素列）を登録する [1]。今回使用した voca ファイルを図 8 に示す。grammar ファイルとは構文制約を単語のカテゴリを終端規則として記述したファイルである。[1] 今回使用した grammar ファイルを図 9 に示す。dfa ファイルとはオートマトン定義ファイルであり，grammar で記述される文法制約を有限個の状態と状態遷移のある振る舞いの抽象的なモデルに変換したものである。[1] 今回使用した dfa ファイルを図 10 に示す。term ファイルとは単語カテゴリ番号と実際のカテゴリ名の対応をさせるために用いるファイルである [2]。今回使用した term ファイルを図 11 に示す。dic ファイルとは単語を認識させるための辞書ファイルである。後述の dict ファイルとの相違点は jconf ファイルに指定する単語を認識させるためである。今回使用した dic ファイルを図 12 に示す。dict ファイルとは単語を認識させるための辞書ファイルである。[3] dic ファイルとの相違点は dfa ファイルと連動させる点にある。今回使用した dict ファイルを図 13 に示す。

1 すみません	すみません
2 吸います すいます	
3 吸いません	すいません
4 吸わないです	すわないます
5 禁煙です	きんえんです
6 喫煙です	きつえんです

Fig. 7 yomi ファイル

```
1 % TABACO
2 すみません      s u m i m a s e n
3 吸いません      s u i m a s e n
4 吸います s u i m a s u
5 吸いけません s u w a n a i d e s u
6 吸わないです k i t s u e n d e s u
7 禁煙です      k i n e n d e s u
8 % NS_B         s l l B
9 <S>
10 %NS_E         s l l E
11 </S>
```

Fig. 8 voca ファイル



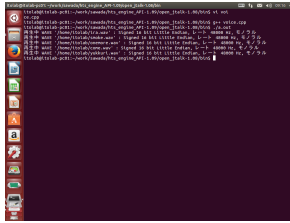


Fig. 4 音声の出力

```

////////////////////
/// Module mode ready
/// waiting client at 10500
////////////////////

```

Fig. 5 サーバモード

```

..... System Information end .....
notice for feature extraction (dl).
=====
* Control room normalization for real-time decoding. *
* NOTICE: The first input may not be recognized, since *
* no initial word is available on startup. *
=====
stat: capture audio at 10000Hz
stat: adin_alsa: latency set to 32 msec (chunk = 512 bytes)
error: adin_alsa: unable to get pcm info from card control
warning: adin_alsa: skip output of detached audio device info
STAT: AD-in thread created
<=====
<=====
<=====
<=====

```

Fig. 6 文字化け

```

1 S      : NS_B TABACO_ NS_E
2 TABACO_ : TABACO|

```

Fig. 9 grammar ファイル



```

1 0 2 1 0 0
2 1 0 2 0 0
3 2 1 3 0 0
4 3 -1 -1 1 0

```

**Fig. 10** dfa ファイル

1	0	TABACO
2	1	NS_B
3	2	NS_E

**Fig. 11** term ファイル

1	すみません	sumimasen
2	吸います	sui masu
3	吸いません	sui masen
4	吸わないです	suwanaidesu
5	禁煙です	kinendesu
6	喫煙です	kit suendesu

**Fig. 12** dic ファイル

10	[すみません]	sumimasen
20	[吸います]	sui masen
30	[吸いません]	sui masu
40	[吸わないです]	suwanaidesu
50	[禁煙です]	kit suendesu
60	[喫煙です]	kinendesu
71	[<s>]	silb
82	[</s>]	silE

**Fig. 13** dict ファイル