

# 目次

第 1 章 序 論	3
1.1 研究の背景	3
1.2 研究の目的	3
1.3 論文の構成	5
第 2 章 飲食店での接客について	7
2.1 声とは	7
2.2 各社音声認識用ソフトの調査	10
2.3 音声認識とは	11
2.4 各社音声合成用ソフトの調査	12
2.5 音声合成とは	13
第 3 章 検討用のシステムについて	18
第 4 章 音声認識実験	21
4.1 実験目的	21
第 5 章 結言	34
5.1 本研究のまとめ	34
5.2 今後の課題	34

参 考 文 献

35

謝 辞

36

# 第1章

## 序論

### 1.1 研究の背景

近年ソフトバンク社がPepperを開発した。また、長崎県に存在するテーマパーク「ハウステンボス」で宿泊する際に利用すると考えられる「変なホテル」の受付には女性タイプと恐竜タイプの接客ロボットや話しかけることで照明のオンオフを行ってくれるロボットなどが配備されている。本研究は飲食店にて接客業務での人件費削減を行うため、ロボットを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開発する必要があると考え、本研究を開始した。飲食店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。図1.1にpepper、図1.2にホテルに存在する女性タイプの接客ロボット、図1.3に恐竜タイプの接客ロボットの外観を示す。

### 1.2 研究の目的

飲食店にて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、マイクを準備し飲食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することを目的とする。図1.4に音声認識をしようとしたときのイメージ図

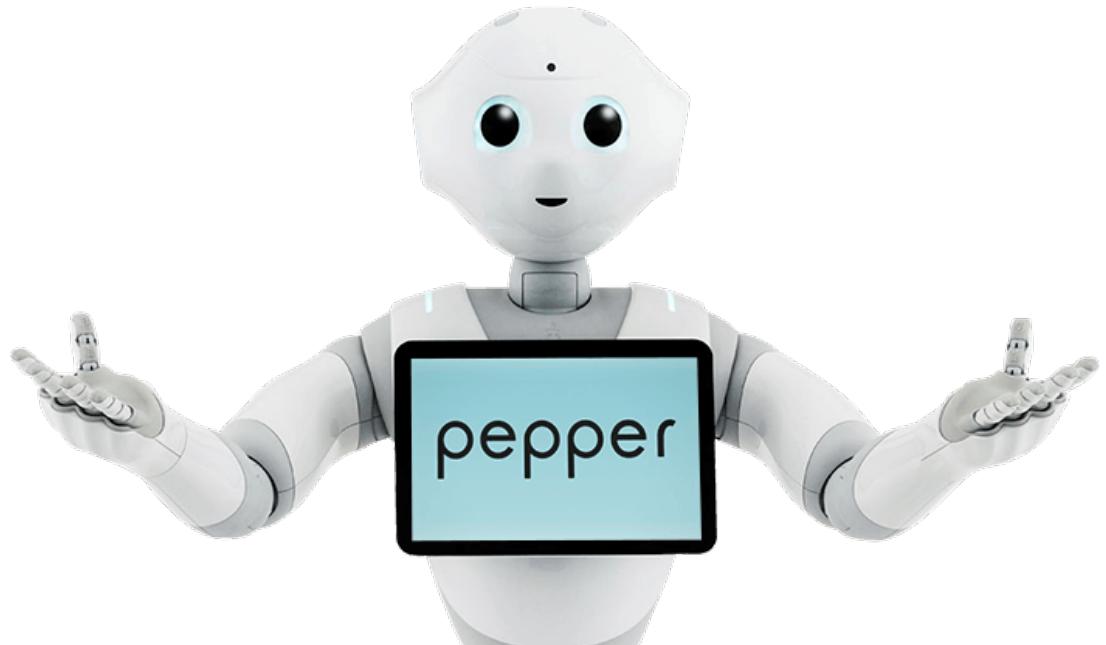


図 1.1 pepper



図 1.2 女性タイプの接客ロボット

を示す。



図 1.3 恐竜タイプの接客ロボット

### 1.3 論文の構成

この論文は以下のように構成されている。

1章では、音や音声について、研究背景について述べた。

2章では、本研究に使用する、音声を認識するための道具及び発生する音声について述べる。

3章では、本研究のために構築した検討用のシステムについて述べる。

4章では、前章で述べた検討用のシステムを使用して行った実験について述べる。

5章では、本研究のまとめや今後の課題について述べる。



図 1.4 様子

## 第2章

# 飲食店での接客について

### 2.0.1 接客の業務内容

お客様から注文を取り、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。その接客の業務内容の流れをフローチャートを図2.1に示す。

### 2.0.2 研究として検討する接客の業務内容

接客の仕事内容をフローチャートとして挙げたが今回、接客をするロボットの開発を行うにあたり、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接客の業務内容となる。

## 2.1 声とは

声とは空気の振動であり波である。つまりsinカーブで表すことができ、人間が声として発する「あ」という音一つから「卒業研究」のような単語、「私の職業は学生であるため勉強をすることが仕事である。」と言った文章を声で発している限りsinカーブで表すことができる。人間の声をマイクに入れてそのデータをの周波数成分を明らかにするため

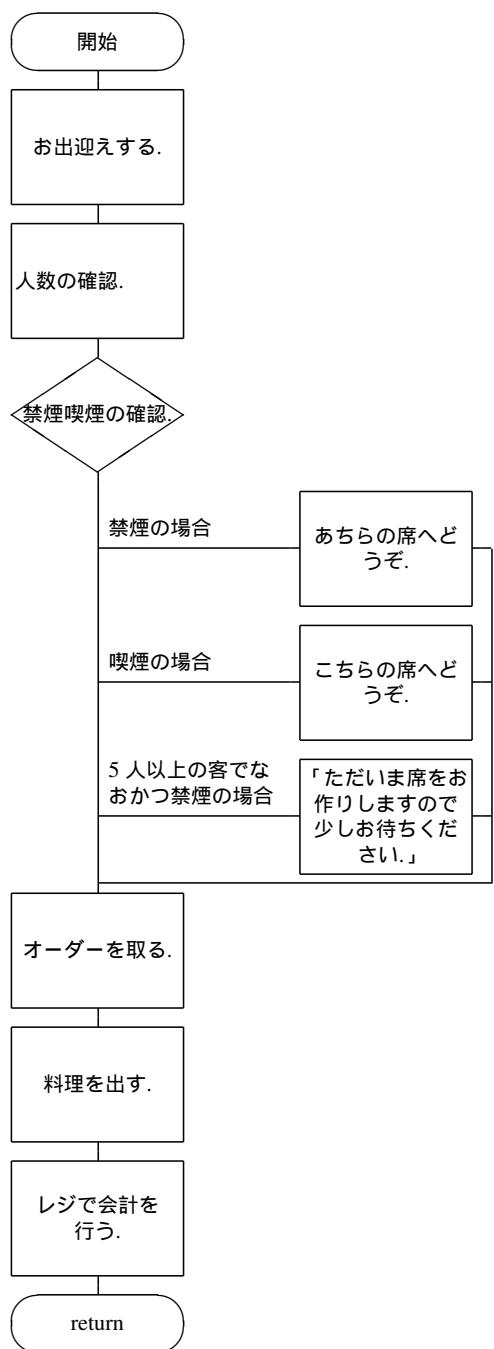


図 2.1 接客方法の流れ

には特別な手段を用いる必要がある。それは音声にフーリエ変換を掛けることである。

### 2.1.1 フーリエ変換とは

まず、フーリエ変換は式 2.1.1 で示すことが出来る。式 2.1.1 より関数  $f(t)$  と  $e^{-j\omega t}$  を掛けあわせて積分することにより 2 つの関数が似ていれば絶対値が大きな値、似ていなければ小さな値を取りますので関数  $f(t)$  は時間の関数でもあるためオシロスコープでも確認することが可能である。また、回転の成分としてスペクトラムアナライザーでも確認することが出来る。<sup>[1]</sup>

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t} dt = 0 \quad (2.1)$$

### 2.1.2 高速フーリエ変換とは

前述の通り声は空気の振動のため実際に視覚的に見ようとした場合、特別な手段を用いなければならない。ここではマイクを使い視覚的に表示することで音声の波形を見ることができると考え、実際に audacity という音声の編集ソフトを使用して音の波形を視覚的に確認することに成功した。ここで、音の波形を高速フーリエ変換にかける。高速フーリエ変換は英語で Fast Fourier Transform と呼ばれていることから頭文字を取って FFTとも呼ばれている(以下 FFT とする)。前述に示したフーリエ変換に各音専用の係数をかけることで FFT を行うことができる。図 2.2 に「あ」の音声波形、図 2.3 に「い」の音声波形、図 2.4 に「う」の音声波形、図 2.5 に「え」の音声波形、図 2.6 に「お」の音声波形、図 2.7 に「あ」に FFT をかけた波形、図 2.8 に「い」に FFT をかけた波形、図 2.9 に「う」に FFT をかけた波形、図 2.10 に「え」に FFT をかけた波形、図 2.11 に「お」に FFT をかけた波形を示す。

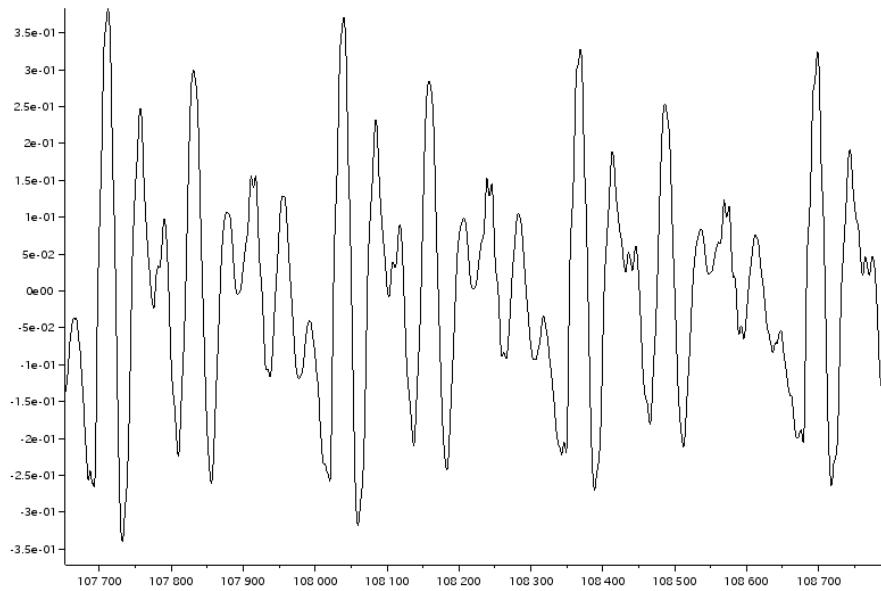


図 2.2 「あ」の音声波形

## 2.2 各社音声認識用ソフトの調査

音声によるロボットを制作するにあたり、まずは音声を認識することが必要条件である。そのため、音声を認識するプログラムを作成し使用することが考えられるのだが時間的コストがかかることが予想された。そのため音声を認識するためのソフトを調査した。調査した結果,Julius, ドラゴンスピーチ,docomo 音声認識 API,Ami Voice,Google Speech API の 5 つが候補として挙がった.Julius を除いた音声認識ソフトは無料のソフトを選択した場合使用期限があり、有料版を選んだ場合金銭的コストがかかるため本研究では linux でも使用でき無料かつ使用期間の制限がない Julius を選択した。表 2.1 に音声認識一覧を示す。また、本研究で使用することとした音声認識ソフト Julius のアイコンを図 2.12 に示す。

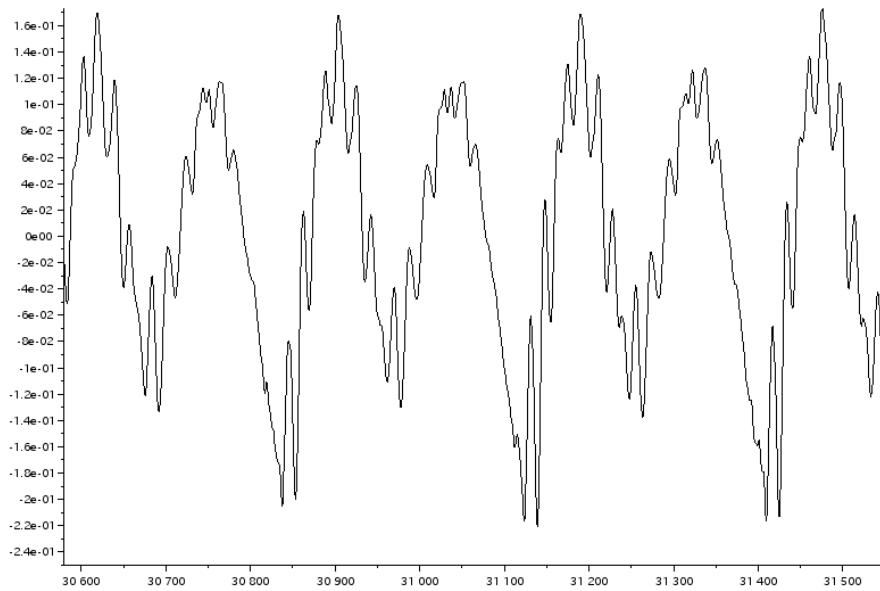


図 2.3 「い」の音声波形

表 2.1 音声認識一覧

ソフト名	無料である	使用期限がない	認識率
Julius			90% 以上
ドラゴンスピーチ	×	×	不明
docomo 音声認識 API	×	×	不明
Ami Voice	×	×	不明
Google Speech API	×	×	不明

## 2.3 音声認識とは

音声認識とは音声をマイクより入力し、入力された音声を文字として表すことを指す。

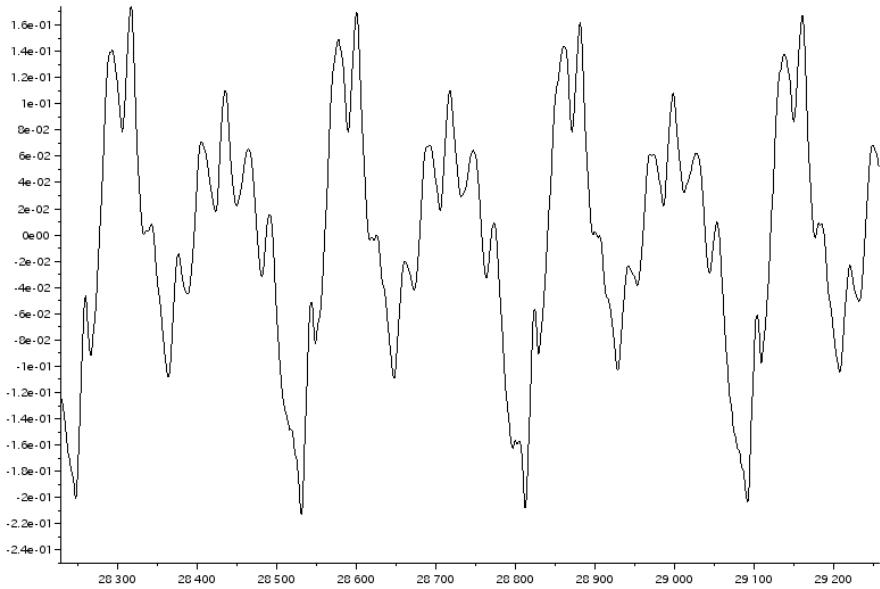


図 2.4 「う」の音声波形

## 2.4 各社音声合成用ソフトの調査

接客口ボットは音声を認識した後に音声を発して店内での案内も行わなければならぬ、よって音声合成ソフトを使用することで音声を作成し言語を用いて发声させることを考えた。そこで音声を作成するためのソフトの調査を行った結果,Open-JTalk,Vocaloid,AITalk,VoiseText,テキストーク,Clipboard Translator,棒読みちゃん,歌声合成ツールUTAUの8つのソフトが候補に挙がった。本研究ではまず音声を发声させることができ、なおかつ音声の速さが変更できるOpen-JTalkを選択した。表2.2に音声合成一覧を示す。また、本研究で使用することとした音声合成ソフトOpen-JTalkのアイコンを図2.13に示す。

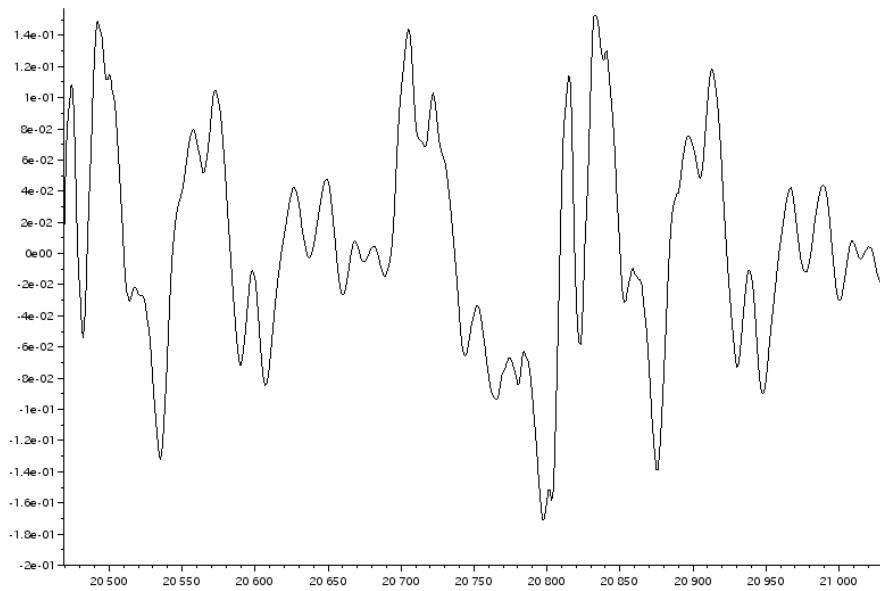


図 2.5 「え」の音声波形

表 2.2 音声合成一覧

ソフト名	音の早さが変更可能	無料である	使用期限がある	男女性の声で話せる	感情表現可能	日本語以外が話せる	備考
Open-JTalk			×			不明	なし
Vocaloid		×		不明	不明	不明	歌声合成ソフト
AITalk	不明	×					海外の言語は 36 種類
VoiceText	不明	×		不明	不明	不明	なし
テキストーク		×			×	不明	Open-JTalk と提携した場合にのみ女性の声が使用可能
ClipBoard Translator	不明	×		不明	不明	不明	Windows のみ使用可能
棒読みちゃん		×		不明	不明	不明	なし
歌声合成ツール UTAU	不明	×		不明	不明	不明	歌声合成ソフト

## 2.5 音声合成とは

音声合成とは人間の音声を人工的に作り上げることを指す。

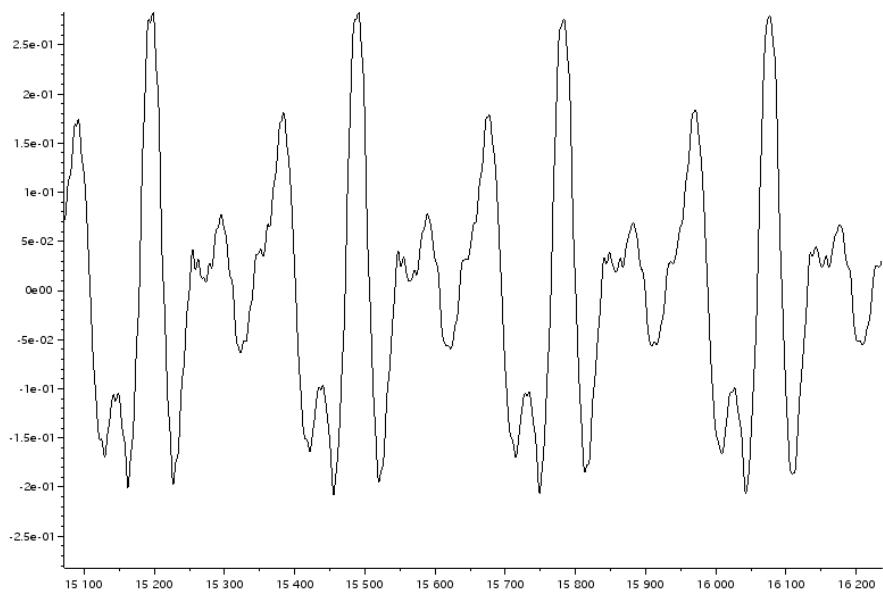


図 2.6 「お」の音声波形

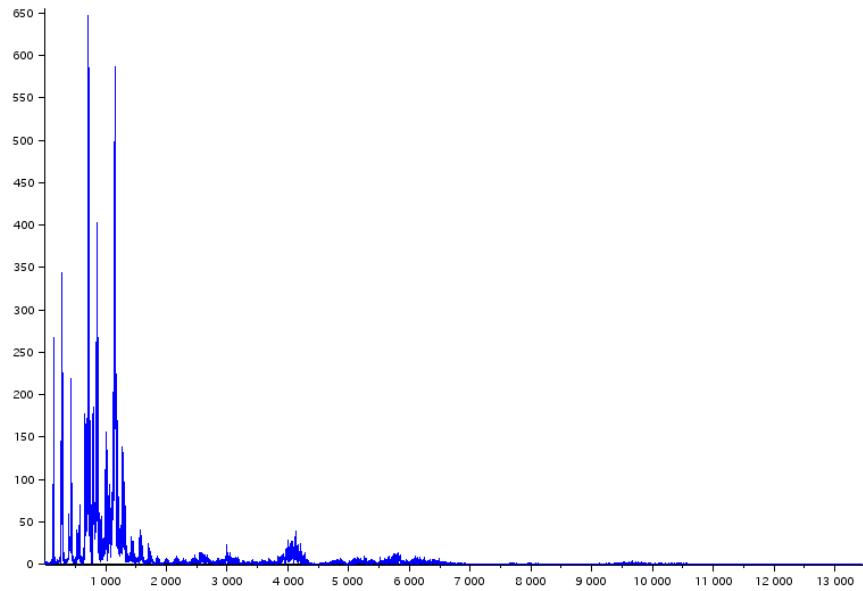


図 2.7 fft を掛けた「あ」の波形

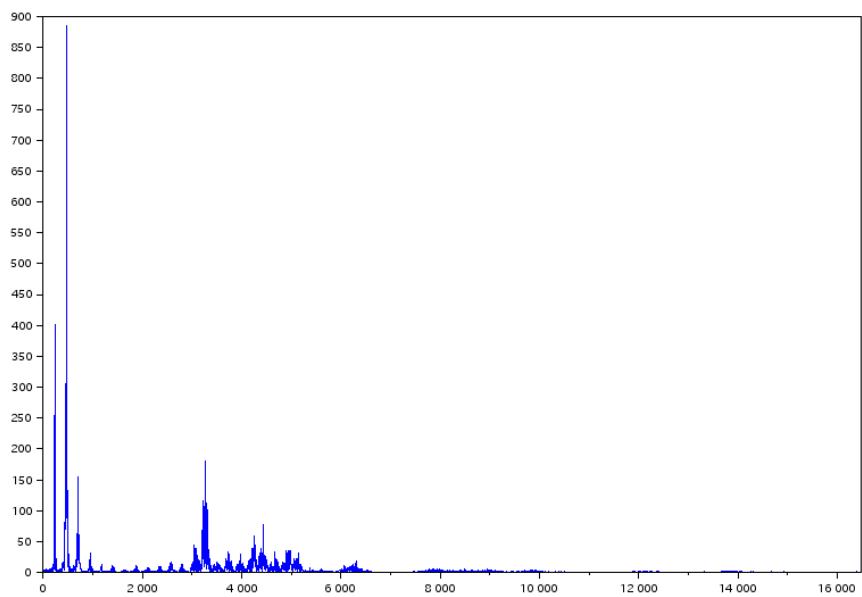


図 2.8 fft を掛けた「い」の波形

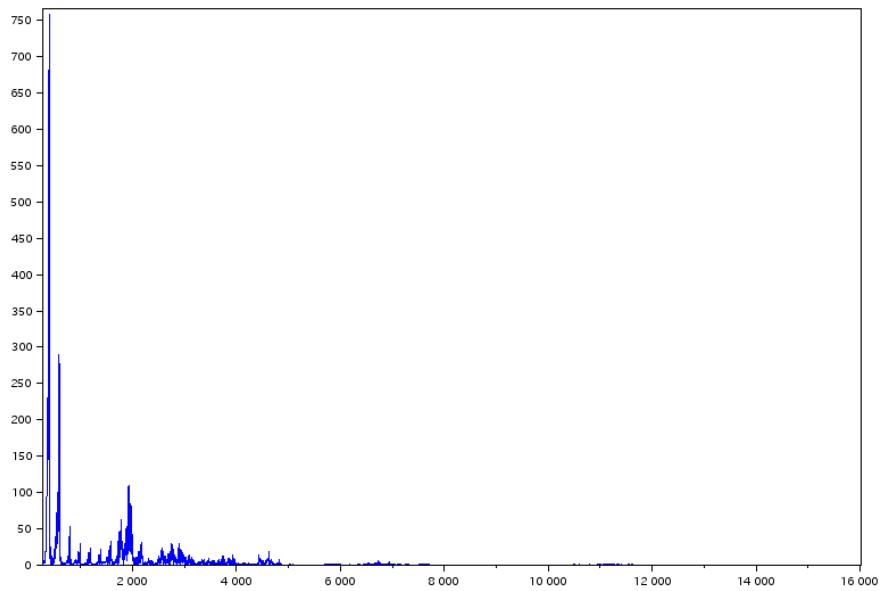


図 2.9 fft を掛けた「う」の波形

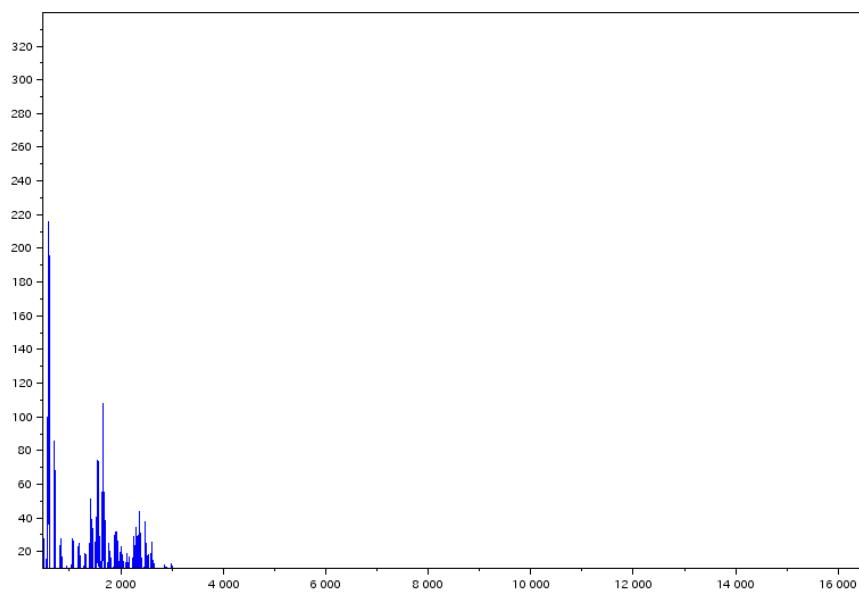


図 2.10 fft を掛けた「え」の波形

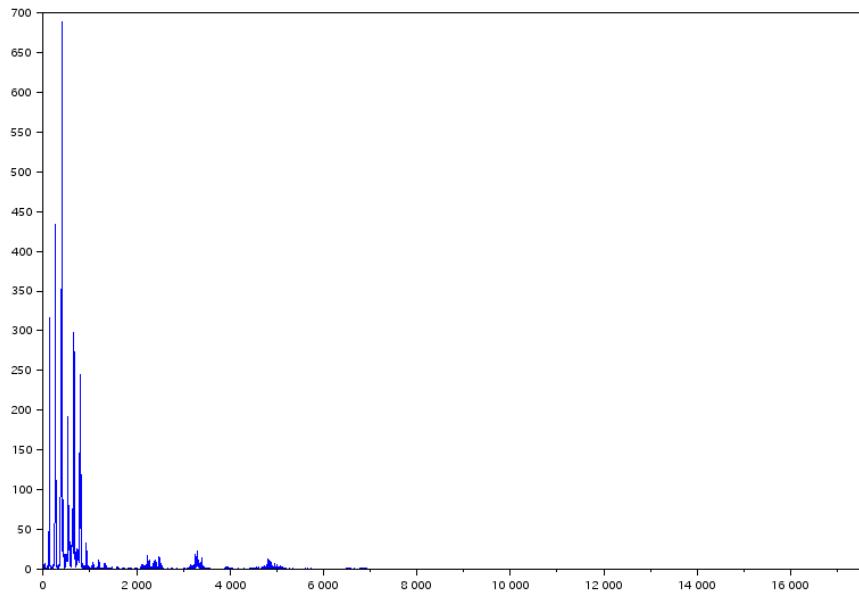


図 2.11 fft を掛けた「お」の波形

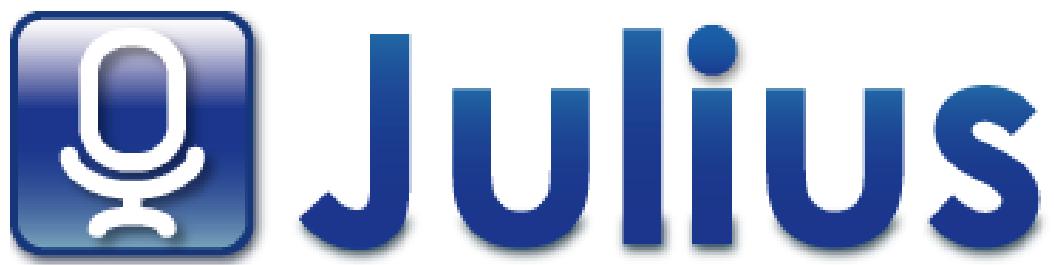


図 2.12 Julius



図 2.13 Open-JTalk

## 第3章

# 検討用のシステムについて

今年度のシステムは図3.1に示すシステムを使用して音声認識を行う。



図3.1 使用したシステム

### 3.0.1 実験で使用するマイク

本研究では buffalo 製のマイク (BSHSM04BK) を使用した。その図を図 3.2 に示す。



図 3.2 使用したマイク

### 3.0.2 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。<sup>[2]</sup> 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で使用でき

るため、今回はサーバモードで起動し、クライアント側で認識した単語に対応した音声を発生させることを目指す。また本研究で使用したJuliusは最新版としてバージョン4.4.2が平成28年9月12日にGithubにて更新されている。しかし、研究を始めた時点でバージョン4.3.1が最新版だったためこれを継続して使用することとした。

#### Juliusの仕組み

Juliusの音声認識アルゴリズムは、ツリートレリス探索方式を基礎とするアルゴリズムである。全体は2パス構成となっており、2段階に分けて認識処理を行う。まず、第1パスでは入力全体に対して荒い認識処理を行い、有望な候補の集合をある程度絞り込む。このとき、簡便なモデルや近似計算を用いることで高速に処理を行う。第2パスでは詳細な認識処理を行うが、その際に第1パスの結果を参照しながら探索を行うことで、必要な部分にだけ精密な再計算を行って、最終的な最尤解を求める。<sup>[3]</sup>

#### 3.0.3 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。ここでは案内のために使用する音声の内容をWAVファイルとして保存するために使用する。「今日はあちらの席へどうぞ」、「こちらの席へどうぞ」の2つをWAVファイルとして保存した。

## 第4章

# 音声認識実験

### 4.1 実験目的

今回 Julius を使用して音声を認識させ, 100 回分の認識結果を集計し平均を取ることを目的とする.

#### 4.1.1 実験方法

マイクを実装したパソコンを使用し, Julius を端末より起動させる. Julius を使用し、「禁煙です」, 「喫煙です」, 「吸います」, 「吸いません」, 「吸わないです」の認識を行う. このとき 100 回分の認識したときと認識しなかったときの結果を集計する. 100 回分の認識結果の平均を式 4.1.1 より算出する.

$$x = a/100 \quad (4.1)$$

#### 4.1.2 実験結果

#### 4.1.3 Julius を用いて行った実験結果

今回 Julius を用いて実験を行い, 前節に示した式より算出した認識率は禁煙が 99 %, 喫煙が 100 %, 吸いますが 100 %, 吸いませんが 30 %, 吸わないです

ですが 100 % という結果が得られた。また、マイクと被験者との距離を遠く取り、実験を行った。そのときの認識率は禁煙が 90 %、喫煙が 86 %、吸いますが 43 %、吸いませんが 69 %、吸わないですが 71 % だった。別条件での認識結果を表 4.1 に禁煙、表 4.2 に喫煙、表 4.3 に吸います、表 4.4 に吸いません、表 4.5 に吸わないですがの認識結果、表 4.6 に認識結果一覧を示す。また、そのときのグラフを図 4.1 に示す。また、マイクとの距離を遠くして認識実験を行った結果、という結果が得られた。そのときの表を表 4.7、表 4.8、表 4.9、表 4.10 表 4.11、に示す。グラフにしたものを作成したものを図 4.2 に示す。

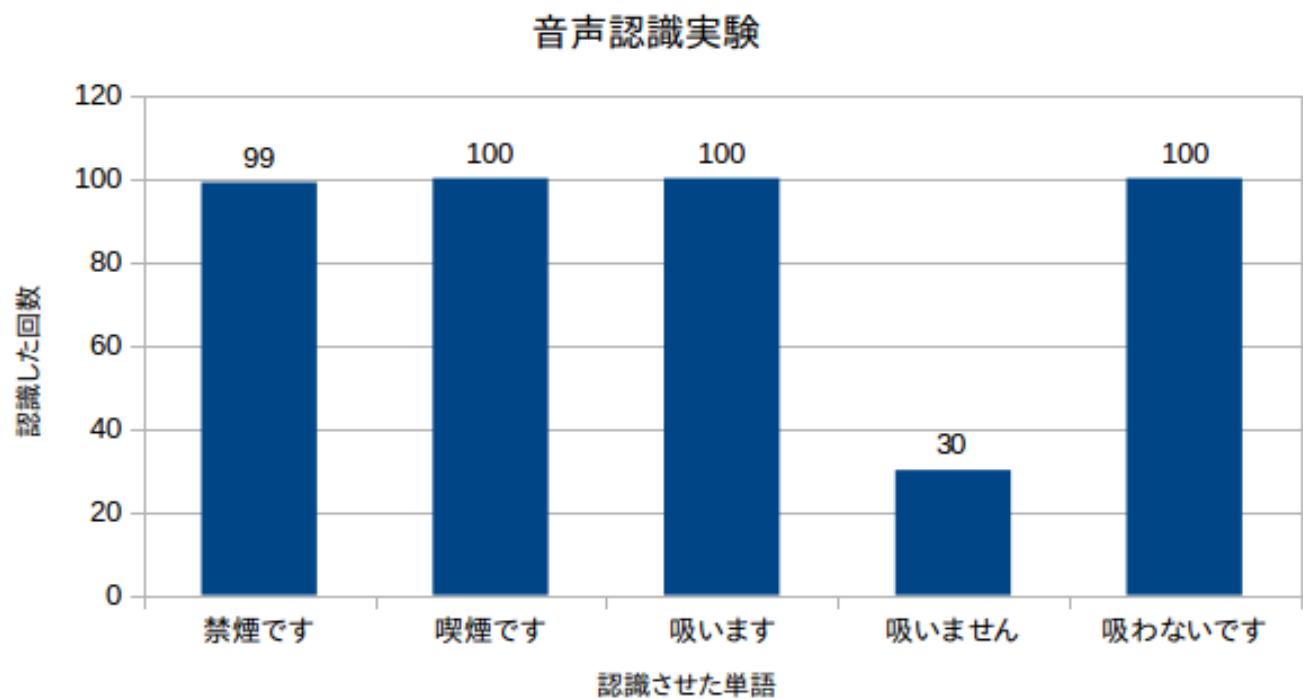


図 4.1 認識結果グラフ

#### 4.1.4 考察

前述の音声認識実験での結果より認識率が 99 %, 30 % の単語が存在するがこれは「禁煙です」が「吸います」、「吸います」が「吸いません」と間  
22

表 4.1 禁煙の認識結果

回数	禁煙です	回数	禁煙です	回数	禁煙です	回数	禁煙です
1		26		51		76	
2		27		52		77	
3		28		53		78	
4		29		54		79	
5		30		55		80	
6		31		56		81	
7		32		57		82	
8		33		58		83	
9		34		59		84	
10		35		60		85	
11		36		61		86	
12		37		62		87	
13		38		63		88	
14		39		64		89	
15		40		65		90	
16		41		66		91	
17		42		67		92	
18		43		68		93	
19		44		69		94	
20		45		70		95	
21		46		71		96	
22		47		72		97	
23		48		73		98	
24		49		74		99	
25		50		75		100	

違えたためである。また、「禁煙です」が「吸います」と誤認識をした原因は不明だが「吸います」と「吸いません」が極めて似ている単語であるため誤認識をしたのではないかと考えられる。また同じ被験者の方に

表 4.2 嘸煙の認識結果

回数	喰煙です	回数	喰煙です	回数	喰煙です	回数	喰煙です
1		26		51		76	
2		27		52		77	
3		28		53		78	
4		29		54		79	
5		30		55		80	
6		31		56		81	
7		32		57		82	
8		33		58		83	
9		34		59		84	
10		35		60		85	
11		36		61		86	
12		37		62		87	
13		38		63		88	
14		39		64		89	
15		40		65		90	
16		41		66		91	
17		42		67		92	
18		43		68		93	
19		44		69		94	
20		45		70		95	
21		46		71		96	
22		47		72		97	
23		48		73		98	
24		49		74		99	
25		50		75		100	

協力していただき、マイクとの距離を変更して行った音声認識実験では「禁煙です」、「喰煙です」、「吸います」、「吸いません」、「吸わないです」の認識が成功した数が減少したのは input rejected by short input という認

表 4.3 吸いますの認識結果

回数	吸います	回数	吸います	回数	吸います	回数	吸います
1		26		51		76	
2		27		52		77	
3		28		53		78	
4		29		54		79	
5		30		55		80	
6		31		56		81	
7		32		57		82	
8		33		58		83	
9		34		59		84	
10		35		60		85	
11		36		61		86	
12		37		62		87	
13		38		63		88	
14		39		64		89	
15		40		65		90	
16		41		66		91	
17		42		67		92	
18		43		68		93	
19		44		69		94	
20		45		70		95	
21		46		71		96	
22		47		72		97	
23		48		73		98	
24		49		74		99	
25		50		75		100	

識不可を示す結果が出たためである。「吸いません」の音声認識実験を行った際に認識不可以以外にも「喫煙です」と認識されたことがあった。

表 4.4 吸いませんの認識結果

回数	吸いません	回数	吸いません	回数	吸いません	回数	吸いません
1		26		51		76	
2		27		52		77	
3		28		53		78	
4		29		54		79	
5		30		55		80	
6		31		56		81	
7		32		57		82	
8		33		58		83	
9		34		59		84	
10		35		60		85	
11		36		61		86	
12		37		62		87	
13		38		63		88	
14		39		64		89	
15		40		65		90	
16		41		66		91	
17		42		67		92	
18		43		68		93	
19		44		69		94	
20		45		70		95	
21		46		71		96	
22		47		72		97	
23		48		73		98	
24		49		74		99	
25		50		75		100	

表 4.5 吸わないですの認識結果

回数	吸わないです	回数	吸かないです	回数	吸わないです	回数	吸わないです
1		26		51		76	
2		27		52		77	
3		28		53		78	
4		29		54		79	
5		30		55		80	
6		31		56		81	
7		32		57		82	
8		33		58		83	
9		34		59		84	
10		35		60		85	
11		36		61		86	
12		37		62		87	
13		38		63		88	
14		39		64		89	
15		40		65		90	
16		41		66		91	
17		42		67		92	
18		43		68		93	
19		44		69		94	
20		45		70		95	
21		46		71		96	
22		47		72		97	
23		48		73		98	
24		49		74		99	
25		50		75		100	

表 4.6 認識結果一覧

単語	禁煙です	喫煙です	吸います	吸いません	吸わないです
計	99	100	100	30	100
平均	0.99	1	1	0.3	1

表 4.7 別条件での禁煙の認識結果

回数	禁煙です	回数	禁煙です	回数	禁煙です	回数	禁煙です
1		26		51		76	×
2		27		52		77	
3	×	28		53		78	
4	×	29		54		79	
5		30		55		80	
6		31		56		81	
7		32		57		82	
8		33		58		83	
9		34		59	×	84	
10		35		60		85	×
11		36		61		86	×
12		37		62		87	×
13		38		63		88	
14		39		64		89	
15		40		65		90	×
16		41		66		91	
17		42		67		92	
18		43		68		93	
19		44		69		94	
20		45		70		95	
21		46		71		96	
22		47		72		97	
23		48		73		98	
24		49		74		99	
25		50		75	×	100	×

表 4.8 別条件での喫煙の認識結果

回数	喫煙です	回数	喫煙です	回数	喫煙です	回数	喫煙です
1		26		51	×	76	
2		27		52	×	77	
3		28		53	×	78	
4		29		54	×	79	
5		30		55		80	
6		31		56		81	
7		32		57		82	
8		33	×	58		83	
9		34	×	59	×	84	
10		35		60		85	
11		36		61		86	
12		37		62		87	
13		38		63		88	
14		39	×	64		89	
15		40		65		90	
16		41		66		91	
17		42		67		92	
18		43		68		93	
19		44		69		94	
20		45		70		95	
21	×	46		71		96	
22		47		72		97	×
23		48		73		98	
24		49	×	74		99	×
25		50	×	75		100	×

表 4.9 別条件での吸いますの認識結果

回数	吸います	回数	吸います	回数	吸います	回数	吸います
1	×	26	×	51	×	76	×
2		27		52		77	×
3	×	28		53	×	78	
4	×	29		54	×	79	
5	×	30	×	55	×	80	
6	×	31		56	×	81	
7	×	32	×	57	×	82	
8	×	33		58		83	
9		34	×	59	×	84	
10	×	35		60	×	85	
11		36	×	61	×	86	
12		37	×	62		87	
13	×	38		63	×	88	
14	×	39		64	×	89	
15		40		65	×	90	×
16	×	41		66	×	91	×
17	×	42		67	×	92	×
18		43	×	68	×	93	×
19		44	×	69		94	×
20		45	×	70	×	95	×
21	×	46		71		96	×
22		47	×	72	×	97	×
23		48	×	73		98	×
24	×	49	×	74	×	99	
25		50	×	75	×	100	

表 4.10 別条件での吸いませんの認識結果

回数	吸いません	回数	吸いません	回数	吸いません	回数	吸いません
1		26		51		76	×
2		27		52	×	77	
3		28		53		78	×
4		29		54		79	
5		30		55		80	×
6		31		56	×	81	×
7	×	32		57		82	×
8		33		58	×	83	
9		34		59		84	×
10		35		60		85	×
11		36		61	×	86	
12		37		62		87	
13		38	×	63	×	88	
14		39		64		89	
15		40	×	65		90	
16	×	41	×	66		91	
17	×	42		67		92	×
18		43	×	68	×	93	
19		44	×	69		94	
20		45		7×		95	
21		46	×	71	×	96	
22	×	47		72	×	97	
23		48		73	×	98	
24	×	49		74	×	99	
25	×	50		75	×	100	

表 4.11 別条件での吸わないですの認識結果

回数	吸わないです	回数	吸かないです	回数	吸わないです	回数	吸わないです
1		26		51		76	×
2		27		52	×	77	
3		28		53	×	78	×
4		29		54	×	79	
5		30		55	×	80	
6		31		56		81	
7		32		57		82	
8		33		58	×	83	
9		34		59		84	
10		35		60		85	×
11		36		61		86	×
12		37		62		87	×
13		38		63	×	88	×
14		39		64	×	89	×
15		40		65		90	×
16		41		66		91	×
17		42		67		92	
18		43		68	×	93	
19		44		69		94	×
20		45	×	70		95	
21		46		71	×	96	×
22		47	×	72		97	×
23		48		73		98	×
24		49	×	74	×	99	
25		50	×	75	×	100	×

表 4.12 別条件での認識結果一覧

単語	禁煙です	喫煙です	吸います	吸いません	吸わないです
計	90	86	43	69	71
平均	0.9	0.86	0.43	0.69	0.71

## 音声認識実験

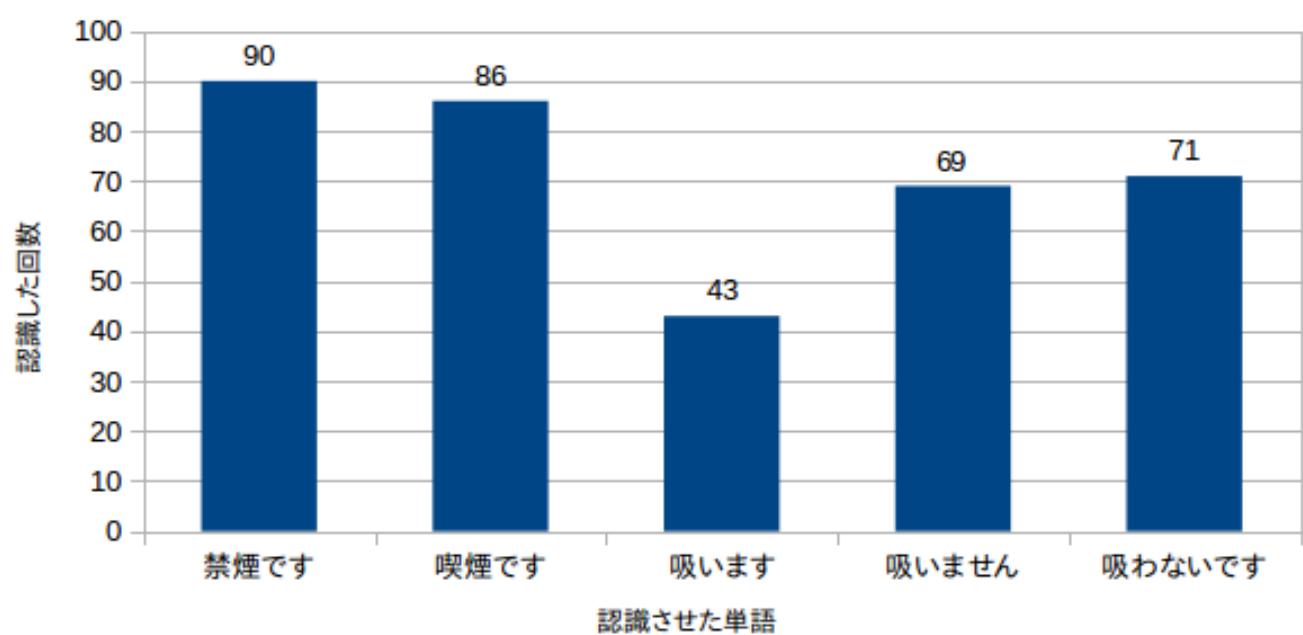


図 4.2 別条件での認識結果グラフ

# 第5章

## 結言

### 5.1 本研究のまとめ

音声を認識する接客口ボットのシステムを構築するにあたり Julius を用いることで音声認識を行うことが可能となった。また、音声の認識結果より平均を算出することで Julius の有用性を確かめることができた。今回は音声を作成するのみとなつたが音声を認識した後に発声させる予定だった音声を WAV ファイルとして作成し、保存することができた。これによって音声を認識した後に対応した音声を発声させることが出来る。

### 5.2 今後の課題

今後は研究者だけでなく研究者以外の人間を実験対象とし、認識率を算出する。また音声を認識したあとに認識した単語に対応した音声を発生させるプログラムを TCP/IP を介して組み合わせることにより、音声を認識し、認識した結果より対応した音声を発声させるためのシステムとして扱うことのできるシステムの確立がなされるのではないかと考える。また、このシステムを用いることで音声認識を用いた接客口ボットの開発を進めることができるのではないかと考える。

# 参考文献

- [1] ディジタルアナログ信号処理のためのやさしいフーリエ変換松尾博著森北出版社 1986年3月31日出版 P4-P5
- [2] Julius book 第7章 言語モデル・<https://julius.osdn.jp/juliusbook/Ja/desc\lm.html>・2016/09/02 閲覧
- [3] Julius book 第8章 認識アルゴリズムとパラメータ [https://julius.osdn.jp/juliusbook/ja/desc\\_search.html?id=2538692](https://julius.osdn.jp/juliusbook/ja/desc_search.html?id=2538692)・2017/01/20 閲覧

# 謝辞

本論文作成にあたりテーマの決定、研究の考え方、方法のまとめ方など全てにおいて長期にわたって厳しくも熱意のあるご指導、ご鞭撻していただいた、伊藤恒平教授に厚く御礼申し上げます。

特に分析においても論文の書き方においても論文を何度も読んでいただき、指導していただいた伊藤恒平教授に大変ご苦労をかけてしましましたことにも心よりお詫び申し上げたいです。

また音声認識実験に協力していただいた寺山森也君に厚く御礼申し上げます。

その他、助けていただいた多くの皆様に心から感謝しております。ありがとうございました。