まえがき

人間は自己のいくつかの感覚器官から得られる情報を統合して世界を経験している。川端康成の有名な小説『雪国』の「国境の長いトンネルを抜けると雪国であった」という一節の中には、トンネルの暗がりから白銀の雪面の上に紺の夜景色が広がり、雪の冷気を感じるという世界に移動したという視覚感覚と皮膚感覚と平衡感覚が融合して体験されている。このいくつかの感覚器官を通して複合的に得られる情報を「マルチモダリティ」という。人はマルチモダリティを利用して環境や他者とインタラクションしているのである。

物理学用語でいう「インタラクション」とは、原子や素粒子が重力や電磁力によって相互に影響し合うことを指す。人間の行為としてのインタラクションは、複数の個人がたがいに働きかけ合うことを指す。マルチモダリティを受け取り、相手に投げ返し合うやり取りを「マルチモーダルインタラクション」という。

人工知能学者の素朴な夢は、人間と対等なマルチモーダルインタラクション が行える人工物をつくることにある。

「ロボットと言えば?」と隣の人に聞いてみる。「鉄腕アトムでしょうか」という答えが返ってきた。本書の究極の目標はみんなを助けてまわる心優しき 『鉄腕アトム』をつくることにあるのかもしれない。

人と身体の組成や動力源は異なるものの、言語音声や身振り手振りを使って同じようなコミュニケーションができる人工知能物をいつの日かつくりあげてみたいと願う。そのためには、人間に本来備わっている音声に乗せて発される言葉、手足や胴体、頭といった身体の動きによって表現される身振りなど、人が人に情報を伝達するために利用してきた中核的なメディアがどのように産出・理解されているのかを知る必要がある。音や身体という複数の発信機を

「マルチモーダルインタラクションメディア」と呼ぼう。本書には、マルチモーダルインタラクションメディアを構成する言語音声や目や手や体といった身体動作を利用するインタラクションシステム生成のための種火を記す。

1章では言語・音声・非言語にかかわるインタラクションシステム生成のためにこれまでなされてきた研究を振り返る。2章ではインタラクティブに言語をやり取りする対話を行うための根幹となる言語処理について取り上げる。人間が送受信する言語の音声音響信号処理は3章で解説する。4章では、視線や身体動作を含むマルチモーダルインタラクションメディアがどのように使われているのかを調べるために、人々の振舞いを収集・分析するための手法を説明する。そして、それをどう分析するのか、最後の5章ではマルチモーダルインタラクションを扱った人と人、人とコンピュータの社会的インタラクションのモデル化に関する研究を紹介する。

マルチモーダルインタラクションメディアにかかわる領域が多岐にわたることから、つぎの3名で分担して執筆した。

飯田仁:1,2章,相川清明:1,3章,榎本美香:1,4,5章

言語に関する研究、音声に関する研究、非言語に関する研究は、この順で歴史的に登場したものであり、読者におかれては1章から順に読み進めていただきたい。そして、人とマルチモーダルインタラクションが可能な知的人工物を創造するための有志となっていただければ幸いである。

2013年8月

著者を代表して 榎本美香

1章 人と人、人とコンピュータのインタラクションを介在するメ	ディア
	2
1.2 言語処理のあゆみ ———————	4
1.3 音声音響処理のあゆみ	5
1.4 身体を持った人工物のあゆみ —————	6
演 習 問 題	
2章 言 語 処 理	
2.1 文 法———————————————————————————————————	
2.1.1 文の基本構造―――――	
2.1.2 文法の基本―――――	
2.1.3 形態素解析—————	
2.1.4 句構造文法—————	
2.2 形 式 言 語——————————————————————————————————	27
2.2.1 文脈自由文法————————————————————————————————————	
2.2.2 構 文 解 析——————————————————————————————————	
2.3 自然言語の統計的・確率的な性質	
2.3.1 単語の出現頻度 ————————————————————————————————————	
2.3.2 Nグラム統計モデル	
2.3.3 確 率 文 法——————————————————————————————————	
2.4 言語の理解	
2.4.1 言語の意味理解 —	
2.4.2 語が担う意味役割	
	50

viii <u></u> 国 次	
2.4.3 言語解釈と言語行為	55
演 習 問 題 —————————————————————————————————	
3章 音声音響信号処理	
3.1 アナログからデジタルへ ————	
3.1.1 本章におけるきまり —————	62
3.1.2 音は波	
3.1.3 音のデジタル化 ————————————————————————————————————	
3.2 時間領域と周波数領域 —————	69
3.2.1 周期的波形——————	
3.2.2 フーリエ展開	
3.2.3 フーリエ展開の複素表現 ――――	
3.2.4 フーリエ変換	
3.2.5 FFT —	
3.2.6 z 変 換 —————————————————————————————————	
3.3 さまざまな音信号 ——————	
3.3.1 周波数変化音————————————————————————————————————	
3.3.2 振幅変化音———————	85
3.4 デジタルフィルタ — — — — — — — — — — — — — — — — — — —	86
3.4.1 時系列信号の演算の特性 —————	86
3.4.2 FIRフィルタ―――	
3.4.3 インパルス応答 ――――	89
3.4.4 IIRフィルタ―――	91
3.4.5 最 小 位 相 —————————————————————————————————	96
3.5 スペクトル分析 —————	
3.5.1 短区間の切り出し ――――	99
3.5.2 振幅スペクトルとパワースペクトル ————	101
3.5.3 対数スペクトル	102
3.5.4 スペクトログラム	104
3.6 音声特有の信号処理	105

		次
3.6.1	線形予測分析 ————	
3.6.2	ケプストラム―――	
3.6.3	ケプストラムによるピッチ抽出 ――――	
3.6.4	LPC ケプストラム	
3.6.5	変形相関関数によるピッチ抽出 ————	
.7 音	声分析と声質	
.8 音	声認識と音声合成の音響処理	
3.8.1	メル周波数軸	
3.8.2	MFCC —	
3.8.3	デルタケプストラム	
3.8.4	音声認識における音響処理の基本 ――――――	
3.8.5	類 似 度 ————	
3.8.6	DP マッチング	
3.8.7	隠れマルコフモデル ――――	
習	問 題	
章 ₹	ルチモーダル情報付きデータベースの作品	龙法
 .1 ਵ		
	ルチモーダル情報とは	
.2 実		
	験室データとフィールドデータをめぐる諸相 ―	
	験室データとフィールドデータをめぐる諸相 ― 言語情報付きデータの作成法 ―――――	
. 3 非 4.3.1	験室データとフィールドデータをめぐる諸相 — 言語情報付きデータの作成法 ———— 3人自由会話 ————	
. 3 非 4.3.1 4.3.2	験室データとフィールドデータをめぐる諸相 ― 言語情報付きデータの作成法 ――― 3人 自 由 会話 ――― 合意形成型多人数会話 ――――	
.3 非 4.3.1 4.3.2 4.3.3	験室データとフィールドデータをめぐる諸相 ― 言語情報付きデータの作成法 ―――― 3人自由会話 ―――― 合意形成型多人数会話 ――――――――――――――――――――――――――――――――――――	
.3 非 4.3.1 4.3.2 4.3.3 4.3.4	験室データとフィールドデータをめぐる諸相 ―言語情報付きデータの作成法 ―3人自由会話 ―合意形成型多人数会話 ―日常活動データ ―世代間協働インタラクションデータ ―	
.3 非 4.3.1 4.3.2 4.3.3 4.3.4 .4 言	験室データとフィールドデータをめぐる諸相 ― 言語情報付きデータの作成法 ― 3人自由会話 ― 合意形成型多人数会話 ― 日常活動データ ― 世代間協働インタラクションデータ ― 語・非言語行動のアノテーション ― ―	
.3 非 4.3.1 4.3.2 4.3.3 4.3.4 .4 言 4.4.1	験室データとフィールドデータをめぐる諸相 ― 言語情報付きデータの作成法 ― 3人自由会話 ― 合意形成型多人数会話 ― 日常活動データ ― 世代間協働インタラクションデータ ― 語・非言語行動のアノテーション ― 言語行動のアノテーション ―	
.3 非 4.3.1 4.3.2 4.3.3 4.3.4 .4 言 4.4.1	験室データとフィールドデータをめぐる諸相 ― 言語情報付きデータの作成法 ― 3人自由会話 ― 合意形成型多人数会話 ― 日常活動データ ― 世代間協働インタラクションデータ ― 語・非言語行動のアノテーション ― 非言語行動のアノテーション ― 非言語行動のアノテーション ―	
.3 非 4.3.1 4.3.2 4.3.3 4.3.4 .4 言 4.4.1 4.4.2 .5 言	験室データとフィールドデータをめぐる諸相 ― 言語情報付きデータの作成法 3人自由会話 ― 合意形成型多人数会話 ― 日常活動データ ― 世代間協働インタラクションデータ ― 語・非言語行動のアノテーション ― 非言語行動のアノテーション ― 非言語行動のアノテーション ― 語・非言語情報のデータベース化 ―	
.3 非 4.3.1 4.3.2 4.3.3 4.3.4 .4 言 4.4.1 4.4.2 .5 言 4.5.1	験室データとフィールドデータをめぐる諸相 ― 言語情報付きデータの作成法 ― 3人自由会話 ― 合意形成型多人数会話 ― 日常活動データ ― 世代間協働インタラクションデータ ― 語・非言語行動のアノテーション ― 非言語行動のアノテーション ― 非言語行動のアノテーション ―	

Χ	I	次
^		ν

油	<u> </u>	閂	駬	171
/欠		167	ルス -	

_	
う章 マルチモーダルインタラクション分析	
5.1 話者交替にかかわる話し手と聞き手たちの視線の向き ——	174
5.1.1 方 法———————————————————————————————————	- 176
5.1.2 分析 1: 一般的傾向 ————————————————————————————————————	- 181
5.1.3 分析 2:変則事例の分析 ————————————————————————————————————	184
5.1.4 議	- 190
5.2 人-人インタラクションと人-CG エージェントインタラクショ	ン
の行為交替規則	193
5.2.1 人と人工物の社会的インタラクションにかかわる先行研究 ―――	194
5.2.2 人間どうしの社会的インタラクションにかかわる先行研究 ―――	198
5.2.3 対話資料————	- 201
5.2.4 非言語行動のラベリング	- 204
5.2.5 分析:人-人,人-エージェントの対話の基礎的特徴 ————	- 205
5.2.6 人対人の対話における行為の配置規則の定式化 —————	- 211
5.2.7 人対エージェントの対話において行為の配置規則は守られるか ―	- 217
5.2.8 ま と め―――	- 222
演 習 問 題 —————————————————————————————————	223
引用・参考文献 ————————————————————————————————————	224
演習問題解答—————	230
索 引————————————————————————————————————	
21	

1

人と人、人とコンピュータのインタ ラクションを介在するメディア

◆ 本章のテーマ

本章では、昔から人が持っているコミュニケーションのためのメディア、音声に乗せて発される言葉、手足や胴体、頭といった身体の動きといったマルチモーダルインタラクションメディアが工学的にどのように解明されてきたかについて概観する。言語処理・音声音響処理・身体を持った人工物に関する工学的歴史を振り返り、そのなかに残された解くべき問題を示す。人と対等にわたり合えるインタラクションシステムを実現するためには、マルチモーダルインタラクションメディアの統合的な知能処理が必要なことを知ることからスタートしよう。

◆本章の構成(キーワード)

- 1.1 人間が持つマルチモーダルインタラクションメディア 音声に乗せて発される言葉、手足や胴体、頭といった身体の動きなど
- 1.2 言語処理のあゆみ 単語置換自動装置 三段階式翻訳装置 対話システム
- 1.3 音声音響処理のあゆみ 隠れマルコフモデル(HMM), 統計分布, 音声データベース
- 1.4 身体を持った人工物のあゆみ 6 足歩行ロボット、環境とのインタラクション、マルチモーダル情報の 理解と産出

◆本章を学ぶと以下の内容をマスターできます

- ☞ マルチモーダルインタラクションの重要性
- 言語処理の工学的あゆみと解くべき問題
- ☞ 音声音響処理の工学的あゆみと解くべき問題
- ☞ 身体を持った人工物の工学的あゆみと解くべき問題

1.1 人間が持つマルチモーダルインタラクションメディア

人は生まれながらにして他者と情報を交換する力を持っている。母親と視線を見交わすことに始まり、母語を習得し、生い育つなかで社会的な立ち居振舞いができるようになる。そして、古代より連綿と、自らが世代で培った叡智を次の世代へと受け継いできた。現在われわれが享受している科学技術の恩恵は、人類が持つこの他者と情報のやり取りができる能力のたまものである。

近年マスメディアやソーシャルメディアなどさまざまな新しい情報伝達の形態が進化してきたが、それらメディアの発信者と受信者はやはり人間であり、間にどのような技術が介在しようと、人と人とのやり取りであるという点には変わりない。音声に乗せて発される言葉、手足や胴体、頭といった身体の動きによって表現される身振りは、原始より人が人へ情報を伝達するために利用してきた中核的なメディアである。この人間が自らの身体のさまざまな器官を使って送受信し情報交換を行うことをマルチモーダルインタラクションと呼ぶ。廊下でしばらく会っていなかった友達と出会った場面を考えてみよう。

- ① 少し遠くから友達を発見して、たがいに目を見交わす。
- ② 声が届く範囲までたがいに近づいたら、手を上げて、「よっ」と声をかける。
- ③「久しぶり」「元気してた?」「うーん、ちょっと…」「どうした?」「大学来てなくて…」など相手の話しぶりを見て、たがいの近況を話す。
- ④「また会おうぜ」と肩をたたいて別れ、それぞれ違う方向に歩き出す。

このちょっとした立ち話を行うにあたっても、相手が声の届く範囲に来た頃合いを見計らって、挨拶の声かけやジェスチャを行い、言葉を交わすなかで、相手の声音や顔色を読み取るといったことをしている。そして、また頃合いを見計らって別れの言葉を交わし、たがいの向かう方向へ歩き出す。われわれは普段たいした負担に感じることもなく、実にさまざまな複数の情報を組み合わ

せて他者とコミュニケーションしているのである。人が身体器官から発する情 報は、言語的情報・パラ言語的情報・非言語的情報に分けることができる。こ の分け方は藤崎^{1),†}によって提起されたもので、言語的情報とは、ことばの記 号的情報。すなわち辞書・統語・意味・談話のレベルで、文字言語によって表 現されるものである。パラ言語的情報とは、話者の意図や態度を表すもので、 おもに音声の韻律的特徴によって表される。非言語的情報とは、話者の個人的 な特徴や年齢、性別、健康状態などの身体的状態に関するものと感情などの心 理的状態に関するものとされる。これは音声を対象に提唱された分類である が、現在では、広義に音声情報全体を指してパラ言語情報、ジェスチャや視 線・身体の向き、他者との身体配置関係など、音声以外で表出される情報につ いて非言語的情報と呼ぶことが多い。人が人とコミュニケーションを行うと き、これらの情報を複合的に用いて、さまざまな情報をたがいに発信し受信す るというキャッチボールのようなやり取りであるインタラクションを発生させ ている。

ところが、バーチャルエージェントや実体を持つロボットなどに同様のイン タラクションをさせようとすると、一体どの情報をいつ用いて総合的に判断す るか、途端に難解な問題に直面することになる。先の友達と遭遇する場面を例 にとろう。

- ① たくさんの人が行き交うなかから、どうして友達であると確信がなくと もまず認識し、自分が相手を認識したことをどのようにして相手に伝え、 相手の反応から確かに友達だと確信したことをどのようにして、相手に伝 えるのか。
- ② 声が届く範囲とは、いったい何メートルか。
- ③ 久しぶりに出会った友人と何を話せばよいのか。
- ④ いつ話し終わればよいのか。どのようにしてたがいに別れて歩き出すのか。

[†] 肩付き数字は巻末の引用・参考文献番号を表す。

ここで提示した友達どうしのやり取りにおいて、気軽な挨拶や近況を伝え合う手短な会話では、その言葉を解釈することで友達双方のコミュニケーションが成り立っていることが理解できる。このような対話の理解のためには、それぞれの発話の解釈と対話全体の解釈である言語の解釈が必要となる。このことは、例えば「久しぶり」と発したとき、休暇が久しぶりであるとか、台風が久しぶりであるとかいろいろな状況しだいで「久しぶり」の解釈は異なる。つまり、発話が生じる状況しだいでその手短な発話ではさまざまな解釈があり得るということになる。その解釈が論理上さまざまであっても、人間の対話ではその解釈にあまり困ることなく対話を続けている。

1.2 言語処理のあゆみ

言語を解釈することについての工学的な取組みには、多くの時間を費やして きた。「久しぶり」という名詞が発せられたことをコンピュータが認識するこ とはあらかじめ用意された辞書を参照すれば容易にできる。科学史に名を残す 偉人であるデカルトやライプニッツは、17世紀にドイツ語とフランス語間の 翻訳辞書を用意した**単語置換自動装置**の作成を試み.商取引などに利用され た。商取引の限られた内容に関して、語の意味が一通りに解釈できる範囲で有 効に使われたと考えられる。20世紀に入って. 1930年代にはフランスとソビ エト連邦(現. ロシア)でそれぞれ独立に、辞書を利用して文を構成する語を 特定し、他の言語の文に置き換える三段階式翻訳装置が特許として認められ た。これらの手法は電子計算機が登場する以前の特許であるものの、現在でも コンピュータ処理において言語を取り扱う根幹の処理手法である構文解析の原 型となっている。この構文解析については2章で説明する。ただ、文を構成す る語を特定し、他の言語の語に置き換える操作を基本とする自動翻訳では、語 の解釈が一意に決まるときに限って翻訳が成功するという問題を抱えている。 そのため、先に示した対話例において、「久しぶり」をどのように解釈すれば よいのかという問題を語を中心にして解決することはできない。このような問

題は、その後の日本語の対話形式による質問応答システムや英語音声の対話形式による航空便予約システムなどにおいても十分な解決は見ず、対話の話題を限定することにより、いろいろな解釈が入り得る余地をなくすことで使用を可能としてきた。そのため、言語情報だけでなく、視覚情報、音声情報、対話がなされる環境ならびに状況に関する情報、さらには社会生活上規範とされる振舞いなどまでを取り込んだ対話の解釈が必要となる。

1.3 音声音響処理のあゆみ

音声情報の認識と合成に関する研究がこれまで盛んに行われている。音声認 識は 1976年の DARPA(Defence Advanced Research Project Agency)の音声 理解プロジェクトの貢献が大きい。1000単語リアルタイム音声認識を課題と したこのプロジェクトではカーネギーメロン大学の HARPY システムのみが課 題を達成した。この目標設定は研究を推進させる強い原動力となった。この時 期から、カーネギーメロン大学、AT&Tベル研究所、BBN、IBM、NTT、東北 大学電気通信研究所、東芝、NEC ほか、多くの大学、家電メーカーが競争に 参入した。例えば、NTT は ANSER システムにより音声対話による銀行の残高 照会を実用化した。実用を想定したマルチモーダル対話としては東芝のハン バーガー販売システム Tosburg がある。飛躍的な音声認識性能向上に貢献した のは IBM の**隠れマルコフモデル**(HMM: hidden markov model) を用いた認 識方式である。1997年に発売された IBM の ViaVoice は大語彙連続音声認識を **具現化したものであり、音声認識の研究はこれで終了したとささやかれたほど** の高性能であった。しかし、それから10年以上の歳月を経た現在でも音声認 識の研究は綿々と続けられている。その理由としては、以下のような現状があ る。現在の音声認識は音響情報、言語情報ともに統計分布を用いた方法であ る。しかし、日常社会には新しい言葉や言い回しが出現し、古い言い回しがす たれていく。これらへの対応は現在の音声認識のきわめて重大な課題である。 子供の声、高齢者の声、方言、外国なまりのある声、非常時の発声の認識も現 在の音声認識の課題である。

一方、音声合成で画期的な性能を示したのは Klatt シンセサイザを用いた DECTalk である。これも 1980 年頃の技術である。この頃の音声合成の主流は 規則合成で、人の発声器官のモデルを複数フィルタで構成したものである。し かし、自然性の面で限界があり、現在の音声合成の主流は波形素片を接続して 音声波形を作成する方法である。このほか、隠れマルコフモデルによる合成も 高い性能を示している。

音声対話システムの発展はデータベースに依存する。米国では TIMID. ATIS などの音声データベースが早期に構築され、MIT の Voyager システムな どの代表的な音声対話システムが提案されてきた。日本でも日本語講演音声 コーパス (Corpus of Spoken Japanese) が構築され、自然発話の認識に貢献し ている。

言語処理と音声認識・音声合成の双方を実装、実用化したものとして Apple の Siri があり、話題を呼んだ。しかし、1 対 1 の質問応答はできるものの、連 続した会話ができるわけではない。それは、いまここでその相手と何を話すべ きかという言語情報や音声情報だけでは解決できない問題を含んでいるからで ある。

身体を持った人工物のあゆみ 1.4

人は「久しぶり」という言語を算出・理解するために、それがどの場所で誰 が誰に向かって発した言葉であるかを前提にしている。この言葉を理解するた めには、まず知的人工物自身がどの環境のなかに埋め込まれているかを認識し なければならない。そもそも、友達とどこかで遭遇するには、ある環境のなか をその人工物が移動している必要がある。環境のなかを移動するためには、ま ず人工物は環境の凹凸や障害物を知覚して、それを乗り越えたり避けたりしな ければならない。この環境とのインタラクションを実装した有名なものに. 1990 年初頭にブルックスの考案した **6 足歩行ロボット**のゲンギスロボットが



図1.1 ブルックスの「ゲンギスロボット」

ある。図1.1に示すように6本の足を持ち、昆虫のような見かけである。

前頭部にならぶ集電センサが赤外線を検出すると、その方向へ障害物を越えて動くように設計されている。3次元空間モデルなどの認知処理を経ず、環境知覚がダイレクトに運動へつながる人工物として当時論争を呼んだ。しかし、地面の上をロバストネスに動き回るこの形態のロボットは火星探索機として採用されている。ゲンギスロボットは地表との実用的なインタラクションが可能なレベルまで進化したのである。ブルックスは、ロボットを以下のように定義している²⁾。

- ・現実の状況下に存在すること(situatedness):抽象的な記述ではなく実世界に組み込まれていて、センサを通して触れる、いまここの世界の現実の状況がその行動に直接影響を与えること。
- ・**物理的な身体を持っていること** (embodiment): 物理的な身体を持ち,外界からの影響をその身体に直接受け、経験することができること。

そして、人とインタラクションが可能なロボットは、人間がその身体の形状を通して得た経験をもとに世界をとらえ抽象化できること、そのロボットの動きが人間にとって社会的な意味を持つものとして扱われることが必要であると述べている。すなわち、マルチモーダルな情報の理解と産出が可能な人工物でなければならないことを意味する。

現在の技術を用いれば、音声入力や視覚入力、触覚入力といった人間が五感 を通して得る外界の知覚が可能になる。音声発信機や眼球、手足といった人間 と類似した身体をロボットに与えれば、社会的な意味を人間が解釈可能な人工物となる。しかし、われわれの周囲を見回しても、人間のサポート役として日常生活に溶け込んでいるロボットは存在しない。早稲田大学の ROBISUKE $^{\dagger 1}$, ATR の Robovie $^{\dagger 2}$, NEC の PaPeRo $^{\dagger 3}$, 奈良先端科学技術大学院大学のたけまるくん $^{\dagger 4}$, 三菱重工のワカマル $^{\dagger 5}$ などが開発されているが、まだ実験室的環境の外には出れておらず、生活のパートナーとしてその存在が不可欠なものとは言いがたい。それは、知覚系から入力された情報を体系的につなぎ合わせ、経験として受け止める知能処理のアルゴリズムが実用化できるほど進化していないからである。さらにいえば、どのような身体動作がどういった社会的意味を持つことになるのか、人間の社会的インタラクションのメカニズムが日常生活をカバーできるほど明らかになっていないからである。

人と対等にわたり合える人工物をつくろうとするならば、つまるところ、人は人とどのようにインタラクションしているのかという行動科学になる。われわれ日本人はいったいどのような言語構造を理解しているのか? われわれが聞く音はどのような波を持ち、どのようにして生成されているのか? 言語や音声以外のコミュニケーションメディアには何があり、どのようにして分析の俎上に乗せればよいのか? それらを統合して行っている社会的インタラクションにはどんな法則性があるのか? 次章からは、こういった疑問に取り組んできた学問の成果を解説する。

演習問題

- [1.1] 言語処理だけでは解けない問題とは何かを書きなさい。
- [1.2] 現在の音声認識技術の課題を書きなさい。
- [1.3] 人とインタラクションする人工物に必要な要素は何かを書きなさい。

^{† 1} http://www.pcl.cs.waseda.ac.jp/study03 02.html

^{† 2} http://www.irc.atr.jp/ptRobovie/robovie-j.html

^{† 3} http://www.nec.co.jp/products/robot/

^{† 4} http://robotics.naist.jp/research/takemaru/

^{† 5} http://www.mhi.co.jp/products/detail/wakamaru.html

索引

1±1		確率文法	38	サイドバンド	85
【あ】		隠れマルコフモデル	5, 126	サウンドスペクト	ログラム
アノテーション	153	下降型解析法	31		104
安定なフィルタ	95	間休止単位	155	残 響	90
[(1)]		完結可能点	174	三段階式翻訳装置	4
		換喩	56	サンプリング間隔	65
移動平均型	88	【き】		サンプリング関数	67
意 味	41, 44			サンプリング周波	数 65
意味記述辞書	54	基音	70	サンプリング定理	67
因果性	96	基本周期	69	サンプル	65
インパルス応答	90	基本周波数	69	[L]	
【う】		共起関係	36		
		極	88	自己回帰型	93
ウィーナー・ヒンラ	/	鋸歯状波	69	自己共分散関数	101
	102	[<]		自己相関関数	102
【え】				修辞構造理論	48
		矩形窓	99	() () () () () () () () () ()	16
エリアシング	67	句構造文法	25	周波数	63
【お】		【け】		周波数変化音	<i>83</i>
	70		20	周波数変調	85
オイラーの公式	72 64	形態素解析	20	周波数変調 主 節	85 16
オイラーの公式 オクターブ	64	形態素解析 形容詞	25	周波数変調 主 節 述語成分	85 16 13, 15, 18
オイラーの公式 オクターブ 音の三要素	64 83	形態素解析 形容詞 ケプストラム	25 109	周波数変調 主 節 述語成分 瞬時角周波数	85 16 13, 15, 18 83
オイラーの公式 オクターブ 音の三要素 折返し誤差	64 83 67	形態素解析 形容詞 ケプストラム 言語行為	25 109 57	周波数変調 主 節 述語成分 瞬時角周波数 瞬時周波数	85 16 13, 15, 18 83 83
オイラーの公式 オクターブ 音の三要素 折返し誤差 音圧レベル	64 83 67 103	形態素解析 形容詞 ケプストラム 言語行為 言語行為論	25 109 57 57	周波数変調 主 節 述語成分 瞬時角周波数 瞬時周波数 上昇型解析法	85 16 13, 15, 18 83 83 31
オイラーの公式 オクターブ 音の三要素 折返し誤差 音圧レベル 音 源	64 83 67 103 116	形態素解析 形容詞 ケプストラム 言語行為 言語行為論 言語行動	25 109 57 57 153	周波数変調 主 節 述語成分 瞬時角周波数 瞬時周波数 上昇型解析法 助動詞	85 16 13, 15, 18 83 83 31 18
オイラーの公式 オクターブ 音の三要素 折返し誤差 音圧レベル 音 源 音声生成	64 83 67 103	形態素解析 形容詞 ケプストラム 言語行為 言語行為論 言語行動 言語の統計モデル	25 109 57 57	周波数変調 主 節 述語成分 瞬時角波数 瞬時周波数 上昇型解析法 助動詞 深層格	85 16 13, 15, 18 83 83 31
オイラーの公式 オクターブ 音の三要素 折返し誤差 音圧レベル 音 源 音声生成 音声データベース	64 83 67 103 116	形態素解析 形容詞 ケプストラム 言語行為 言語行為論 言語行動	25 109 57 57 153	周波数変調主 節 述語成分 瞬時角周波数 瞬時周波数 上昇型解析法 助動詞 深層格 振幅変調	85 16 13, 15, 18 83 83 31 18 52
オイラーの公式 オクターブ 音の三要素 折返し誤差 音圧レベル 音 源 音声生成	64 83 67 103 116	形態素解析 形容詞 ケプストラム 言語行為 言語行為論 言語行動 言語の統計モデル	25 109 57 57 153	周波数変調 主 節 述語成分 瞬時角波数 瞬時周波数 上昇型解析法 助動詞 深層格	85 16 13, 15, 18 83 83 31 18 52
オイラーの公式 オクターブ 音の三要素 折返し誤差 音圧レベル 音 源 音声生成 音声データベース	64 83 67 103 116	形態素解析 形容詞 ケプストラム 言語行為 言語行為論 言語行動 言語の統計モデル	25 109 57 57 153 36	周波数変調主 節 述語成分 瞬時角周波数 瞬時周波数 上昇型解析法 助動詞 深層格 振幅変調	85 16 13, 15, 18 83 83 31 18 52
オイラーの公式 オクターブ 音の三要素 折返し誤差 音圧レベル 音 源 音声生成 音声データベース 【か】	64 83 67 103 116 116	形態素解析 形容詞 ケプストラム 言語行為 言語行為論 言語行動 言語の統計モデル 【こ】 高域通過フィルタ	25 109 57 57 153 36	周波数変調 主 節 述語成分 瞬時角波数 瞬時周波数 上昇型解析法 助動詞 深層格 振幅変調	85 16 13, 15, 18 83 31 18 52 85
オイラーの公式 オクターブ 音の三要素 折返し誤差 音圧レベル 音 源 音声生成 音声データベース 【か】 概念階層	64 83 67 103 116 116 6	形態素解析形容詞ケプストラム言語行為言語行為言語行動言語の統計モデル【こ】高域通過フィルタ高域通機大解析	25 109 57 57 153 36	周波数変調 主 節 述語成分 瞬時角波数 降時周波数 上昇型解析法 助動詞 深層格 振幅変調 【せ】 正規化角周波数	85 16 13, 15, 18 83 83 31 18 52 85
オイラーの公式 オクターブ 音の三要素 折返し誤差 音圧レベル 音 源 音声生成 音声データベース 【か】 概念階層 書換え規則	64 83 67 103 116 116 6	形態素解析 形容詞 ケプストラム 言語行為 言語行為論 言語行動 言語の統計モデル 【こ】 高域通過フィルタ 高調波	25 109 57 57 153 36	周波数変調 主 節 述語成分 瞬時周波数 降時周波数 上昇型解析法 助動詞 深層格 振幅変調 【せ】 正規化角周波数	85 16 13, 15, 18 83 83 31 18 52 85
オイラーの公式 オクターブ 音の三要素 折返し誤差 音圧レベル 音 声生成 音声手がータベース 【 か 】 概念階層 書換え規則 格関係	64 83 67 103 116 116 6	形態素解析形容詞ケプストラム言語行為言語行為言語行動言語の統計モデル【こ】高域通過フィルタ高域通機大解析	25 109 57 57 153 36	周波数変調 主 節 述語成分 瞬時周波数 上昇型詞 光層整 上界型詞 深層格 振幅変調 【せ】 正規化角周波数 正規化 正規 正規 正成数	85 16 13, 15, 18 83 83 31 18 52 85
オイラーの公式 オクターブ 音の三しに 音が返した 音が 音が 音が 音が 音が 音が 音が 音が 音が で る で る で る で の で の で の で の で の で の で り で り の で り の で り の く の く の く の く の く の く の く の く の く の	64 83 67 103 116 116 6 42 15 51 51	形態素解析 形容詞 ケプストラム 言語行為 言語行為論 言語の統計モデル 【こ】 高域通過フィルタ 高調波 構文解析	25 109 57 57 153 36 86 70 31	周波数変調 主 が が が が が が が が が が が が が が が が が が が	85 16 13, 15, 18 83 83 31 18 52 85 82 92 63 25

零点全音	87 64	【な】		[ほ]	
全極型	92	ナイキスト	67	方形波 68	9
線形予測	105	1,41		方形窓 99	9
線形予測係数	105	【は】		ボトムアップ解析法 3	1
【そ】		倍 音	70	ホルマント 108	3
1,51		発 散	95	【ま】	
側波帯	85	ハニング窓	99	101	
素性	42	ハミング窓	99	窓関数 99	
【た】		パワースペクトル	101	マルコフ線図 3	
	22	パワースペクトル密原		マルチモダリティ 50	3
帯域消去フィルタ	86	半音	64	マルチモーダルインタラク	^
帯域通過フィルタ ターン構成単位	86 156	搬送波	85	ション 2,50)
単語辞書データ	25	【ひ】		【め】	
単語置換自動装置	4	非言語行動	153	名詞 15, 25	5
単語の出現頻度	35	ビブラート	84	名詞節 13, 15, 17, 18, 19, 25	
単語の長さ	36	比喩表現	56		
単 文	16	表層格	51	(ゆ)	
【ち】		標本化関数	67	ユール-ウォーカー方程式	
[5]		品 詞	15	108	3
調音	117	[ئد]		[0]	
直交関数	71				
陳述表示型	58	不安定なフィルタ	95	量子化 68	Э
【て】		フィードバック	91	(れ)	
低域通過フィルタ	86	フィルタ 負帰還	86 92	 レビンソン-ダービンのアル	
でシベル	103	副詞節	16, 19	ゴリズム 108	R
データ窓	99	複文	16, 16	連接可能性行列 24	_
テプリッツ型行列	108	ブラックマン窓	99	連体節 17, 18	
伝達関数	87	フーリエ展開	71	,_,,,,,	
141		文 18	5, 19, 20	【ろ】	
(と)		文法要素	15	6 足歩行ロボット	6
	15, 16, 18	[^]		【わ】	
投 射 トップダウン解析法	157	並列節	17, 20	話者移行適格場 174, 19	a
トップタリン解例伝	: 31 86	変調周波数	17, 20 84	両有移行適恰場 174, 195 話者交替システム 174	
1. 1. 1. 1.	66		04	in a 文省	
			ı	/ PAP / P	r

$\hspace{1cm} \diamondsuit \hspace{1cm} \hspace{1cm} \diamondsuit$				
[A]		(H)		[P · R]
AM	85	HMM	5, 126	PCM 69
AR	93	Hz (ヘルツ)	63	RST 48
[D]		[I·T]		$(\mathbf{S} \cdot \mathbf{T})$
DP マッチング	126	IIR フィルタ	94	sinc 関数 <i>67</i>
DTW	123, 126	LPC 分析	105	TCU 156
(F)		$[\mathbf{M} \cdot \mathbf{N}]$		TRP 199
FIR フィルタ	88	MA	88	【数字】
FM	85	MFCC	119	2-グラム 36
FM音	83	<i>N</i> -グラムモデル	36	

── 著 者 略 歴 ──

榎本 美香(えのもと みか)

- 1995年 神戸大学教育学部初等教育学科卒業
- 1998年 奈良女子大学大学院文学研究科修士課程修了(教育学専攻)
- 2000年 株式会社エイ・ティー・アール知能映像通信研究所研修研究員
- 2004年 千葉大学学術研究支援員
- 2006年 東京農工大学産学官連携研究員
- 2007年 千葉大学大学院自然科学研究科博士後期課程修了(情報科学専攻)博士(学術)
- 2007年 東京工科大学嘱託研究員
- 2008年 東京工科大学助教
- 2012年 東京工科大学講師

現在に至る

飯田 仁(いいだ ひとし)

- 1972年 早稲田大学理工学部数学科卒業
- 1974年 早稲田大学大学院理工学研究科修士課程修了(数学専攻)
- 1974年 日本電信電話公社基礎研究部勤務
- 1986年 国際電気通信基礎技術研究所出向
- 1998年 ソニー株式会社コンピュータサイエンス研究所勤務
- 2000年 博士(工学)(東京工業大学)
- 2002年 東京工科大学教授 現在に至る

相川 清明(あいかわ きよあき)

- 1975年 東京大学工学部電気電子工学科卒業
- 1980年 東京大学大学院工学系研究科博士課程修了(電子工学専攻) 工学博士
- 1980年 日本電信電話公社武蔵野電気通信研究所入所
- 1989年 カーネギーメロン大学コンピュータサイエンス学部客員研究員
- 1990 年 NTT ヒューマンインタフェース研究所勤務
- 1992 年 国際電気通信基礎技術研究所人間情報通信研究所勤務
- 1996 年 NTT ヒューマンインタフェース研究所勤務
- 1999 年 NTT コミュニケーション科学基礎研究所勤務
- 2003年 東京工科大学教授 現在に至る

マルチモーダルインタラクション

Multimodal Interaction

© Enomoto, Iida, Aikawa 2013

2013年10月17日 初版第1刷発行

*

検印省略

著 者 榎 本 美 香 飯 \mathbb{H} 相 Ш 清 明 株式会社 発行者 コロナ社 牛来真也 代表者 印刷所 萩原印刷株式会社

112-0011 東京都文京区千石 4-46-10

発行所 株式会社 コ ロ ナ 社

CORONA PUBLISHING CO., LTD.

Tokyo Japan

振替 00140-8-14844 · 電話 (03) 3941-3131(代)

ホームページ http://www.coronasha.co.jp

ISBN 978-4-339-02784-6

(安達) (製本:愛千製本所)

Printed in Japan



本書のコピー、スキャン、デジタル化等の 無断複製・転載は著作権法上での例外を除 き禁じられております。購入者以外の第三 者による本書の電子データ化及び電子書籍 化は、いかなる場合も認めておりません。

落丁・乱丁本はお取替えいたします