

まえがき

筆者が、情報処理の観点から音声の研究を始めたのは、課程博士号を取得し、東京大学に任官することとなった1977年4月からであり、以来、40年弱、一貫して音声研究を進めてきている。しばらくの間、学部の講義には、音声を中心テーマとした講義はなかったが、音声認識や音声合成が一般にもなじみのある技術となったのを期に、音声にテーマを絞った学部3年生向けの講義として、「言語・音声情報処理」が、2008年度からスタートした。

講義のスタートに先駆けて、本書の執筆依頼を受けたが、忙しさにかまけて先延ばしするうちに、音声合成への隠れマルコフモデルの導入など、新しい内容が次々と加わり、なかなか構想がまとまらなかった。コロナ社の編集スタッフの励ましで、各章ごとに講義の内容を拡充し、ようやく完成することができた。その間にも、隠れ層が複数の多層ニューラルネットワークを効果的に学習する枠組み(deep learning)が構築され、一時、下火となっていたニューラルネットワークを音声認識に利用する試みが盛んに行われるなどしている。実際、音声関係の主要国際会議であるIEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) や International Speech Communication Association (ISCA) の INTERSPEECH などで、多くの研究論文を見ることができる。この流れは、更に、声質変換などの音声合成分野にも波及してきている。筆者の研究室でも、研究テーマとしているが、まだ、評価や手法が定まっていない面もあり、学部学生を主たる読者と想定している本書では特に言及しなかった。筆者が研究をスタートした頃は、動的なパターン照合が音声認識の主要技術として用いられていたのを考えると隔世の感がする。

音声信号・情報処理の研究が本格的に行われるようになってから、70年を超える歳月が経過し、音声認識や音声合成は当たり前の技術として、我々の生活に入り込むようになっていく。当初は、発話の制約が多く、認識誤りで使いにくかった音声認識も、雑音下で、ある程度自由な発話をしてでも高精度で認識できるようになっている。音声関係の会議では、音声認識、音声合成、言語翻訳の融合技術である音声翻訳を用いて、異言語間の音声対話のデモがよく行われるが、その品質の高さには脅かされる。ただ、このような機器を用いて、自由に海外で生活できる日が来るまでには、まだ、だいぶ時間がかかるような気がしている。それは、このような機器が、多量のデータベースを用いた統計的処理に、支えられたものであり、人間の音声言語活動に対する理解が必ずしも進んでいないことにある。

ビッグデータを用いることで、単なる量の拡大から、質の向上が得られることは論を待た

ないが、音声によって伝達される内容を適切に理解したり、自由な発話スタイルの音声を生成したりするためには、人間の音声生成、受容過程に関する息の長い研究が不可欠であると考えている。統計的手法の導入により、音声合成や音声認識は、これから学ぶ学生にとって、昔と比べると、ある意味、敷居が低くなっているが、音声の生成などの音声研究を進める上での基礎も、ぜひ忘れずに学んでいただきたい。

最後に、本書の執筆に際し、忍耐強く励ましていただいた、コロナ社の関係各位に謝意を表す。

2015年3月

広瀬啓吉

目 次

1. 序 論

2. 音声と情報伝達

2.1 文字言語と音声言語	4
談話室 フィラー	4
2.2 音声の特徴	5
2.3 音声によるコミュニケーション	7
談話室 音声対話システム	7
本章のまとめ	8
理解度の確認	8

3. 音声生成とモデル

3.1 発音器官と音声の生成	10
3.2 音（オン）と音素	11
3.3 音声の生成過程と周波数特性	14
3.4 音 源	16
談話室 基本周波数とピッチ（周波数）	18
3.5 声道伝達特性	19
3.5.1 波動方程式と一般解	19
3.5.2 均一音響管	20
3.5.3 不均一音響管	21
3.5.4 子音の伝達特性と反共振	22
3.5.5 電気回路との対応	23

3.5.6 一般の1次元音響管	25
3.6 放射特性	26
3.7 調音結合	27
3.8 韻律的特徴	28
談話室 臨界制動2次線形系	29
本章のまとめ	30
理解度の確認	30

4. 音 声 分 析

4.1 窓 掛	32
4.2 離散信号化	33
4.3 短時間エネルギーと短時間自己相関関数	34
談話室 窓掛と時間領域の処理	35
4.4 周波数スペクトル	36
談話室 スペクトログラム	38
4.5 線形予測分析	40
談話室 線形予測	40
4.6 自己相関法とPARCOR分析	44
4.7 極/フォルマンントの抽出	46
4.8 ケプストラム	49
談話室 メル尺度	52
4.9 基本周波数の抽出	54
4.10 STRAIGHT分析	55
本章のまとめ	55
理解度の確認	56

5. 自然言語処理

5.1 自然言語の解析	58
5.2 形態素解析	58
談話室 文 節	59

5.3	構文解析	60
5.4	意味解析	64
5.5	文脈解析・談話解析	67
談話室	SHRDLU	68
5.6	機械翻訳	69
談話室	文生成	70
	本章のまとめ	72
	理解度の確認	72

6. 音 声 合 成

6.1	テキストからの音声合成	74
6.2	言語処理（文解析）	74
6.3	音韻処理	75
6.3.1	分節的特徴	75
6.3.2	韻律的特徴	76
6.4	音響処理	78
6.4.1	音声波形の生成手法	78
6.4.2	コーパスベース音声合成と波形編集方式	79
6.4.3	ターミナルアナログ音声合成	80
6.4.4	韻律的特徴の合成	81
6.5	HMM 音声合成	84
6.6	柔軟な音声合成：種々の声質・発話スタイルの合成	86
6.7	声質変換	87
談話室	原音声と目標音声の結合ベクトルによる声質変換	89
6.8	概念からの音声合成	90
談話室	音声を造る	90
	本章のまとめ	91
	理解度の確認	92

7. 音 声 認 識

7.1	処理の流れ	94
7.2	特 徴 量	94
	談話室 韻律的特徴と音声認識	96
7.3	LPC ケプストラム距離 (パターン間の距離)	96
7.4	動的計画法による単語照合	97
7.5	統計的決定理論	99
7.6	音響モデル-隠れマルコフモデル	101
	7.6.1 隠れマルコフモデル	101
	7.6.2 前向き確率と後ろ向き確率	104
	7.6.3 観測系列に対するモデルの尤度の評価	104
	7.6.4 状態系列の推定	106
	7.6.5 HMM パラメータの最尤推定	106
	7.6.6 出力確率分布の共通化	108
7.7	言語モデル	109
7.8	探 索	111
	談話室 連続音声認識システムの性能評価指標	112
7.9	頑健な音声認識	112
	談話室 変換行列の特徴	114
	本章のまとめ	115
	理解度の確認	116
	引用・参考文献	117
	理解度の確認；解説	125
	索 引	127

1

序 論

音声による情報の伝達は、人間の社会活動の最も自然かつ重要な基盤となっている。話し手は、発信したい情報がある場合、まず脳内で発声内容を構築する。この際、情報伝達の前提となる共通知識は発話内容に含めないなどの情報を適切に伝えるための戦略を取る。このため、対話の場合など、話し手・聞き手の知識内容や周囲の状況によって発話内容が異なってくる。次に、調音器官に動作指令を出し、発話内容に対応する音声を生成する。音声は音波として空気中を伝搬し、聞き手に到達する。聴覚系により一定の周波数分析が行われ、聞き手の脳内で認識され、当初の発話内容が理解される。この過程は、通信の送信（符号化）、伝搬、受信（復号化）にアナロジーされる。伝搬中の情報の欠落や、符号化・復号化のプロセスの異なりがあると、意思の疎通が障害されることになる。

音声の生成、伝達、知覚の過程を解明するには、脳科学、生理学、音響物理学、音韻論、言語理論などの学問が必要であり、更に、音声の分析、処理には、MRIなどの発声器官の動きの観測手法、音響信号の解析の前提としてのデジタル信号処理、音声認識や合成のコーパスベース手法のための統計理論や機械学習といった学問分野が関係する。このように、音声の研究は、言語学、医学、物理学、工学にわたる学際性の強いものとなっている。したがって、音声に関連した項目は、生成、知覚、分析、符号化、合成、認識、対話、教育、医療と多岐にわたり、それらをすべて紹介することは本書の範囲を超える。本書では、音声応用の基盤技術である音声合成、音声認識を中心に、それらを理解するために必要な音声生成、音声分析に言及するにとどめる。本書で割愛した音声知覚（聴覚）については、巻末文献1)などに、全般的に詳しく解説されているので参照されたい。ヒトの聴覚の機構に関し、長年にわたってモデル化が進められてきた。これに関しては、巻末文献2)などでまとめられている。

アナログの音声信号をデジタル信号とする際に、音声の特徴を利用し、大幅なデータ圧縮

を行うことができる。これを音声符号化と呼び、音声の高効率伝送技術などとして幅広く利用されている。当然、音声分析と関係が深い³⁾が、音声信号と言語情報の対応関係を中心とした本書では、特に取り扱わなかった。通信の重要な分野であり、音声符号化に焦点を当てた書籍も多いので、参考にされたい³⁾†。

本書では、まず、2章では、音声の特徴とそれにより伝達される情報について概観し、次に3章では、音声の生成を説明する。母音、鼻子音、有声・無声破裂音、有声・無声摩擦音を初めとして、音声は特徴が大きく異なる種々の音（オン）から構成される。これを概観した上で、母音を中心にその音響的特徴量との対応を解説する。音声はマイクロフォンを介して電気信号（音声波形）に変換され、フーリエ変換を初めとする種々の信号処理手法を用いて分析が行われる。4章では、音声波形の（デジタル）信号処理手法を紹介する。6章、7章では、音声対話システム、発音教育システムなどの音声応用システムの基盤技術としての音声合成、音声認識を紹介する。近年の進展が著しい分野であり、最先端の研究が行われている。本書では、基本技術を中心に述べる。

音声で伝達される主要な情報は言語情報であり、自然言語処理との関連が深い。この観点から、5章を自然言語処理の概略の説明にあてている。ただ、2章で述べるように、自然言語処理の対象は、主として書き言葉であり、話し言葉である音声言語とは異なる点があることに留意する必要がある。更に、韻律に代表されるように、文字には通常表記されない情報が音声には含まれる。

† 肩付き数字は、巻末の引用・参考文献の番号を表す。

2

音声と情報伝達

人間は、視覚、聴覚を介して、多量の言語情報を効率的に授受している。言語情報は文字あるいは音声として伝達され、前者は文字言語、後者は音声言語と呼ばれる。文字言語は記録性に優れており、多量の文書が保存されている。インターネットの登場により、文字による多量の情報が容易に得られるようになり、その有効な利用方が盛んに研究されている。一方、音声言語は文字言語の発生のはるか以前から人間の情報伝達手段として活用されており、根源的かつ効率的な情報伝達手段である。現代でも、識字率が50%に満たない国が依然存在しているのに対し、音声言語は全員が共有している。

人間のコミュニケーションでは、文字言語あるいは音声言語の表層的な情報に加え、いわゆる言外の情報というものが存在する。音声には、意図や感情といったこのような情報が特に豊富である。本章では、音声によって伝達される情報を整理し、それが、音声のどのような特徴によって伝達されるかを整理する。

2.1 文字言語と音声言語

言語の伝達媒体には、文字と音声があり、それによって伝達される言語をそれぞれ文字言語 (written language)、音声言語 (spoken language) と呼ぶ。書き言葉、話し言葉とも呼ばれるが、音声合成技術やツイッターの登場により、両者の定義には曖昧な部分も存在するようになった。例えば、書かれた文書を音声合成によって読み上げれば、情報は音声によって伝達されることになるが、話し言葉と呼ぶのは適さない。また、ツイッターは文字表記されるが話し言葉であろう。この観点からは、(韻律を抜きにすれば) 文体から両者を議論するのが自然であろう。

文字言語には、公式文書、新聞、教科書、小説、随筆、メモなど、いろいろバラエティーがあるが、基本的には文法に沿ったものである。一方、音声言語にも、ニュース、講義、講演、(目的のある) 会話、雑談などのバラエティーがあるが、文法に沿わない場合が多く存在するようになる。ニュースの場合は、多くの場合、文法に沿った発声が行われるが、講義や講演では、言い誤り、言い淀み、言い直し、倒置、省略など、従来の書き言葉を対象とした文法には沿わない、いわゆる不規則発話が多く存在するようになる。会話や雑談では、相手の発話によって、自身の発話内容が決まるため、文字表記を音声化するというプロセスを経ない自発発話の側面が強くなり、不規則発話の割合が増加する。このような、不規則発話の存在が、音声認識を困難なものとする一つの大きな要因になっている。

文字言語には、印刷と手書きの二つの表記がある。当然、後者の方がバラエティーに富み、その文字認識は困難になる。音声言語には印刷表記はないが、ニュースのように明確に調音される場合と、雑談のようにそうでない場合がある。不規則発話が多くなれば、一般的に調音も明確でないことが多く、(発話ごとの) 変動も大きいため、音声認識をより困難なものとしている。

☞ 談 話 室 ☞

フィラー 不規則発話として、“アー”、“エー”などを挿入するフィラー (filler) がある。発話を伴った休止ということで filled pause とも呼ばれる。従来は、音声の理解に不要なものとして考えられていたが、フィラーの挿入により、後続の句の理解を促進

するという研究が報告されるようになって¹⁾。フィラーは、自発発話を自然なものとする役割を果たすという側面もあり、単に不要なものではない。音声認識での活用も期待されるが、現状では不要なものとして扱われる。

2.2 音声の特徴

3章で述べるように、音声には、個々の音を特徴付ける特徴に加え、音の高さの変化の様子といった特徴が存在し、情報の伝達を担っている。前者は分節的特徴 (segmental feature) と呼ばれ、おもに音声のスペクトルの包絡といった声道形状が担う音素や音節の情報として現れるのに対し、後者は超分節的特徴 (supra-segmental feature) と呼ばれ、アクセント、抑揚、リズムなどの長時間にわたる情報の伝達に重要な役割を果たす。おもに基本周波数、パワー、長さといった音源に関する特徴が情報を担い、韻律的特徴 (prosodic feature) と呼ばれる。分節的特徴は、単語の同定に主要な役割を果たしているのに対し、超分節的・韻律的特徴は、アクセント型として単語の同定を補佐するとともに、統語境界や文の意味、更には話題の焦点といった高次の言語情報の表現に重要な役割を果たしている。更には、態度、意図、感情といった文字言語では直接表現されないパラ・非言語情報の表現に主要な役割を果たしている²⁾。この観点から、分節的特徴は、音声言語のうち、その文字表記に直接対応する情報を担うのに対し、超分節的・韻律的特徴は、文字表記には直接現れない情報を担っているといえる。表 2.1 は、超分節的・韻律的特徴によって伝達される情報を整理して示した

表 2.1 超分節的・韻律的特徴によって伝達される情報

情報の種類	内容
言語情報	アクセント型、声調 統語境界、係り受け 平叙・疑問文 話題・焦点、段落
パラ・非言語情報	態度、意図、感情、個人性

注) 言語情報は文字として直接的に表記される情報であり、パラ・非言語情報は、表記されない情報である。パラ言語情報は、意図や態度のように、言語情報を補完するもの、非言語情報は、個人性のようにそれ以外の情報である。感情をパラ言語情報、非言語情報のいずれに分類するかについては議論がある。

ものである。なお、中国語の四声などの声調は、音節単位で定義され、語義と深く関わるものであるが、前後の音節の声調の影響を受け、句中での位置で物理的特徴が大きく変化するなど、超分節的特徴としての側面も強い。

音源の特徴として、基本周波数、パワー、長さのほかに、音源波形の形状があり、スペクトルの全体の形状が変化する。これは、個人性の伝達などに重要な役割を果たす韻律的特徴である。なお、音源の特徴は個々の音の生成の前提でもあり、特に、日本語の長母音や促音のように、長さが音素の主要な特徴となっている場合には、音素の同定に關与するものとして分節的特徴として捉えるのが適切であろう。

アクセント、抑揚、リズムは、基本周波数、パワー、長さなどの絶対的な値というよりも、その相対的、時間的变化によって表現される。例えば、基本周波数の高さは、男性、女性といった個人性を表現するが、一時点の高さではなく、平均的な高さである。アクセント型や統語境界といった言語情報の伝達は、時間変化としての基本周波数パターンが担う。長さは平均値からのずれが、早い、遅いといった情報を表し、焦点や態度、あるいは意図の伝達に寄与する。このため、現在の統計ベースの音声認識、あるいは音声合成で行われている短時間フレームでの韻律的特徴の取り扱いは、必ずしも適切なものではない[†]。

音声の研究を進める上で、適宜のラベル表記により言語情報などとの対応を明確にした音声コーパスが重要である。分節的特徴については、基本的には、音素あるいは音(オン)をラベリングすることになるが、韻律的特徴は、文字には直接表記されないという問題がある。英語について開発された tones and break indices (ToBI) がよく知られており³⁾、日本語を初めとした各言語へも適用されている。これは、音声の基本周波数パターンなどの韻律的特徴を見ながら、アクセント核、アクセント句境界などのラベル付けを人間が行うものである。ラベル付けの自動化の試みもなされているが誤りも多い。そもそも、ラベル付けを行うラベラーにより、同じ基準でラベル付けが行われているという保証もなく、ラベリングとして確立したものではない。

[†] 声道形状に対応するスペクトル包絡は分節的特徴を表現するものであるが、例えば、意図や感情などが異なれば、その形状は大きく変化する。この観点からはスペクトル包絡の“基準値からのずれ”も韻律的特徴として取り扱うのが適切と考えられる。実際、感情識別などはスペクトル包絡の違いに着目して行われることが多い。

2.3 音声による コミュニケーション

音声によるコミュニケーションには、講演のように、情報の流れが、基本的に1方向の場合と、双方向の場合がある。後者は、対話あるいは会話と呼ばれ、特に2名の場合を対話、それ以上の場合を会話と呼び区別することもある。前後の情報との関連を持ち、1文だけ抜き出しても、省略や照応[†]のために正確な理解が困難な場合が多い。これを文脈 (context) という。特に、対話や会話の場合、相手の発話内容が文脈となるため、その正確な把握が、対話のスムーズな進行に重要である。また、文書の場合と異なり、聞き誤りや忘却によって、対話の破綻に結びつく。更に、文脈には、一連の発話に出てくる情報だけでなく、情報の発信者と受信者の共通知識もあり、これが一致せずに対話の破綻につながることもある。

情報システムから音声によって情報を得ることを目的として、観光案内、番組案内、航空券・宿舎予約、文献検索など、多くの音声対話システムが開発されている。ヒューマノイドロボットでも音声対話が重要である。このような機械との音声対話には、人間どうしの音声対話とは異なる面が多い。人間は相手が機械であると分かると、人間相手とは異なる発音をするのが一般的である。このため、機械と人間の対話の音声コーパス作成は、音声対話システムの構築に重要であるが、これを人間どうしの対話から得ることは難しい。初期の音声対話システムでは、対話内容を拡張すると、音声認識誤りや発話内容理解の誤りが多く発生し、対話が破綻した。このため、音声認識などの音声対話システムの機能の一部を(利用者が分からないように)代行することが行われた。これを Wizard of OZ システムと呼ぶ。

☎ 談 話 室 ☎

音声対話システム ユーザの発話を入力とし、音声認識 文理解 対話管理 文生成 音声合成の過程を経て、応答音声をユーザに提示するシステムである。1990年代になって音声認識が一定の性能を得るようになると、音声対話システムの研究が盛んに行われた。初期のシステムで有名なのは、米国 MIT の Victor Zue らが開発した VOYAGER であろう。画面を見ながらケンブリッジ市内の案内を行うものである。その後、航空便の座席情報を提示する PEGASAS などのシステムを統合し、電話での利用を可能とし

† “それ”などで先行発話の情報を受けること。

た GALAXY が開発された。更に、マルチモーダルシステムとして、計算機上の擬人化エージェントあるいはロボットとの対話に関する研究が進んだ。道案内、レストラン案内、文献検索など、情報検索をタスクとしたシステムが多いが、発話の際の顎や舌の動きを表示する発音教育システムの開発も行われた。長らく研究開発の域を出なかったが、現在では、スマートフォンを利用した Siri やしゃべってコンシェルなど、広く利用されるようになった。

文理解 対話管理 文生成の過程は自然言語処理と深く関係し、人間のような自由な対話を実現するのは容易ではない。このため、もっぱら入力と応答のコーパスから適切なものを選択することを行っている。人間同士の対話では、「寒いですね」というだけで、窓を閉めるという動作に結びつくが、機械でそれを実現するのは至難の業である。一見、賢そうであるが、人間の対話にはまだまだ遠く及ばない。

言語の異なる人間の対話をサポートするシステムとして音声自動翻訳システムがある。文理解 対話管理 文生成の代わりに翻訳を行うが、例えば、主語を省略することが多い日本語からそうではない英語に翻訳する際を考えればわかるように、1 文のみを対象とした翻訳では十分ではなく、対話履歴の管理も重要である。

本章のまとめ

3 章以降の準備として、音声言語と文字言語について説明したのち、音声の特徴とそれによって伝達される情報を整理した。更に、音声コミュニケーションの特徴について、音声対話システムを念頭に置いて概説した。

理解度の確認

問 2.1 文の統語情報は音声のどのような情報によって表されるかについて述べよ。

【あ】

アクセント核 76
 アクセント型 76
 アクセント結合 77
 アクセント指令 81

【い】

異音 14
 異音化 76
 意味解析 65
 咽頭 11
 韻律的特徴 5, 76

【う】

後ろ向き確率 104
 後ろ向き予測残差 45
 運動性言語中枢 10

【え】

エリアシング 34

【お】

音 11
 音響モデル 100
 音響尤度 100
 音声記号 12
 音声強調 112
 音声言語 4
 音声対話 7
 音声符号化 2
 音声翻訳 71
 音素 11

【か】

外界照応 67
 概念からの音声合成 90
 係り受け解析 64
 書換え規則 60
 格構造 65
 格フレーム 65
 格文法 65
 確率文脈自由文法 63
 隠れマルコフモデル 101

【き】

機械翻訳 69
 記述長最小化原理 85
 規則合成 78
 基本周期 18, 50
 基本周波数 18, 54
 基本周波数パターン 28
 基本周波数パターン生成過程
 モデル 28, 81
 共振 19
 狭帯域スペクトログラム 38
 共分散法 42
 極 80
 距離尺度 96

【く】

矩形窓 32
 句構造文法 60

【け】

形式言語 60
 形態素解析 58
 ケプストラム 49
 言語重み 100
 言語モデル 100
 言語尤度 100

【こ】

口腔 11
 後舌母音 21
 高速フーリエ変換 37
 広帯域スペクトログラム 39
 喉頭 10, 11
 構文解析 60
 国際音声記号 12
 コーバスペース音声合成 79
 混合ガウス分布 102
 コンテキストラベル 85

【さ】

最長一致法 59
 残差 40
 残差駆動 79

【し】

自己回帰移動平均モデル 41
 自己相関関数 54
 自己相関法 42
 シソーラス 65
 ジッタ 18
 シマー 18
 重回帰分析 83
 修辞構造解析 67
 終端記号 60
 周波数スペクトル 36
 照応 67
 照応解析 67
 照応詞 67
 状態出力型 101
 指令-応答モデル 27
 深層格 65
 振幅スペクトル 36

【す】

スペクトル包絡 44
 スペクトログラフ 37
 スペクトログラム 37

【せ】

正規化角周波数 34
 正規文法 61
 声質変換 87
 生成文法 60
 声帯 10
 声帯音源 11, 18
 声帯体積流速 17
 声道 11
 声道アナログ方式 78
 声道伝達特性 14
 声門 11
 接近音 14
 接続コスト 79
 接続コスト最小法 60
 折衷型音声合成器 81
 ゼロ代名詞 67
 遷移出力型 101
 線形予測 40
 線形予測係数 40

線形予測符号化 40
 先行詞 67
 線スペクトル 15, 37
 線スペクトル対 46
 前舌母音 21
 選択コスト 79
 選択制限 66

【そ】

相互情報量 66
 側音 14
 ソースフィルタモデル 16, 78

【た】

ターゲットコスト 79
 ターミナルアナログ音声合成 80
 ターミナルアナログ方式 78
 単語直接方式 69
 短時間エネルギー 34
 短時間自己相関関数 34
 端点フリー DP 99
 談話解析 67

【ち】

チャート法 62
 中間言語方式 69
 中心化理論 68
 調音 11
 調音位置 12
 調音器官 11
 調音結合 11, 27
 調音点 12
 調音様式 12
 超分節の特徴 5
 直列接続型 80
 チョムスキーの階層 60
 チョムスキー標準形 61

【て】

定名詞 67
 テキスト音声変換 74
 テキストからの音声合成 74
 テストセット・
 パープレキシティー 110

【と】

統計的機械翻訳 71
 動的計画法 97
 特殊拍音素 12
 トップダウン解析手法 61
 トランスファ方式 69
 トレリス 105

【に】

入力パターン 97

【は】

パープレキシティー 110
 波形選択音声合成 80
 波形選択合成 86
 波形編集方式 79
 破擦音 14
 はじき音 14
 発声器官 10
 ハミング窓 32
 パラ・非言語情報 5
 パラレルコーパス 87
 破裂音 11, 13
 パワースペクトル 36
 反共振 19, 23
 半連続 HMM 103

【ひ】

鼻音 13
 鼻音化 76
 鼻腔 11
 非終端記号 60
 ビタビアルゴリズム 106
 ビタビ経路 106
 ピッチ 18
 ピッチ抽出 18
 標準パターン 97
 表層格 65
 標本化 33
 標本化定理 33
 品詞接続表 59

【ふ】

フィルラー 4
 フォルマント 15, 46
 フォルマント音声合成 80
 フォルマント周波数 21
 複素ケプストラム 49
 フーリエ変換 32
 ふるえ音 14
 フレーズ指令 81
 フレーム 95
 フレーム周期 33
 フレーム長 33
 分割数最小法 59
 文生成 70
 分析合成方式 78
 文節 59
 文節数最小法 59
 分節の特徴 5, 75
 分布定数回路 23

文脈 7
 文脈依存文法 61
 文脈解析 67
 文脈自由文法 61

【へ】

平均振幅差関数 54
 閉鎖音 13
 並列接続型 81
 ベクトル量子化 87, 102
 変形短時間自己相関関数 35
 偏自己相関係数 46

【ほ】

方形窓 32
 放射特性 14
 ボトムアップ解析手法 61
 翻訳モデル 71

【ま】

マイクロプロソディ 29
 前向き確率 104
 前向き予測残差 45
 摩擦音 14
 マルチテンプレート方式 99
 マルチパルス 79

【む】

無声音 11
 無声化 76
 無声破裂音 11, 13
 無声摩擦音 14

【め】

メルケプストラム係数 95
 メル尺度 50, 52
 メル周波数ケプストラム係数 50

【も】

文字言語 4

【ゆ】

有声音 11
 有声音源 11
 有声破裂音 11, 13
 有声摩擦音 14

【よ】

用例に基づく機械翻訳 71

【ら】

乱流音源 18

【リ】

リアルタイム処理 97
 離散フーリエ変換 36
 離散 HMM 101

量子化 33
 臨界制動 2 次線形系 29

【れ】

零点 80

連結学習 108
 連続 HMM 101
 連濁 75

【A】

A* 探索 111
 A-b-S 47
 ARMA model 41

【B】

back-off smoothing 110
 Bakis 型 101
 Baum-Welch 法 107
 beam 探索 111
 bigram モデル 109
 breadth-first 探索 111

【C】

CATR 80
 cepstral mean normalization
 113
 CYK 法 62

【D】

depth-first 探索 111
 DP 照合法 87, 98

【E】

ergodic model 101

【F】

FFT ケプストラム 53, 95

【G】

GMM 88, 102
 grapheme-to-phoneme conver-
 sion 75

【H】

hidden semi-Markov model
 103

HMM 101
 HMM 音声合成 84
 HSMM 86

【I】

island-driven 探索 111

【K】

Kelly 形声道モデル 26

【L】

left-to-right 型 101
 left-to-right 探索 111
 LPC ケプストラム 53
 LPC ケプストラム距離 97
 LPC ケプストラム係数 95
 LPC 分析 40
 LSP 46
 LSP 分析 46

【M】

MAP 推定 114
 Mealy 型 101
 MFCC 50
 MLR 113
 mono-phone model 108
 Moore 型 101
 MSD-HMM 85

【N】

n-gram モデル 60, 109

【P】

PARCOR 係数 46
 PARCOR 分析 46

【S】

SHRDLU 68

SPLICE 113
 STRAIGHT 55, 86

【T】

TD-PSOLA 79
 tied arc 108
 tied mixture HMM 103
 ToBI 6, 81
 trigram モデル 109
 tri-phone model 108

【U】

unigram モデル 109

【V】

VOCODER 55
 Voder 78
 VOT 13

【W】

Wizard of OZ システム 7

【ギリシャ文字】

Δ ケプストラム 95
 Δ パラメータ 85
 Δ^2 ケプストラム 96
 Δ^2 パラメータ 85

【数 字】

0 型文法 61
 1 型文法 61
 1/2 波長共振 21
 1/4 波長共振 21
 2 型文法 61
 2 質量モデル 17
 2 段 DP 照合 99
 3 型文法 61

— 著者略歴 —

広瀬 啓吉 (ひろせ けいきち)

1977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 (東京大学)

2015年 東京大学名誉教授

音声・言語処理

Spoken Language Processing

© 一般社団法人 電子情報通信学会 2015

2015年 5月 18日 初版第1刷発行

検印省略

編者 一般社団法人
電子情報通信学会
<http://www.ieice.org/>
著者 広瀬 啓吉
発行者 株式会社 コロナ社
代表者 牛来真也

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 Printed in Japan

振替 00140-8-14844・電話(03)3941-3131(代)

<http://www.coronasha.co.jp>

ISBN 978-4-339-01842-4

印刷：三美印刷 / 製本：愛千製本所



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
き禁じられています。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