

日本音響学会 編

音響学講座

6

音 声 (上)

滝口 哲也

編著

有木	康雄	鎗木	時彦
戸田	智基	南角	吉彦
藤本	雅清	木下	慶介

共著



コロナ社

音響学講座編集委員会

編集委員長

安藤彰男（富山大学）

編集委員

秋山いわき（同志社大学）	岩野 公司（東京都市大学）
及川 靖広（早稲田大学）	阪上 公博（神戸大学）
坂本 眞一（滋賀県立大学）	佐藤 史明（千葉工業大学）
滝口 哲也（神戸大学）	菖木 禎史（千葉工業大学）
寺澤 洋子（筑波大学）	古川 茂人（日本電信電話株式会社）
松尾 行雄（東北学院大学）	山田 真司（金沢工業大学）
山本 貢平（小林理学研究所）	渡辺 好章（同志社大学）

（2018年12月現在，五十音順）

「音響学講座」発刊にあたって

音響学は、本来物理学の一分野であり、17世紀にはその最先端の学問分野であった。その後、物理学の主流は量子論や宇宙論などに移り、音響学は、広い裾野を持つ分野に変貌していった。音は人間にとって身近な現象であるため、心理的な側面からも音の研究が行われて、現代の音響学に至っている。さらに、近年の計算機関連技術の進展は、音響学にも多くの影響を及ぼした。日本音響学会は、1977年以来、音響工学講座全8巻を刊行し、わが国の音響学の発展に貢献してきたが、近年の急速な技術革新や分野の拡大に対しては、必ずしも追従できていない。このような状況を鑑み、音響学講座全10巻を新たに刊行するものである。

さて、音響学に関する国際的な学会活動を概観すれば、音響学の物理／心理的な側面で活発な活動を行っているのは、米国音響学会（Acoustical Society of America）であろう。しかしながら、同学会では、信号処理関係の技術ではどちらかという手薄であり、この分野はIEEEが担っている。また、録音再生の分野では、Audio Engineering Societyが活発に活動している。このように、国際的には、複数の学会が分担して音響学を支えている状況である。これに対し、日本音響学会は、単独で音響学全般を扱う特別な学会である。言い換えれば、音響学全体を俯瞰し、これらを体系的に記述する書籍の発行は、日本音響学会ならでの活動ということができよう。

本講座を編集するにあたり、いくつか留意した点がある。前述のとおり本講座は10巻で構成したが、このうち最初の9巻は、教科書として利用できるよう、ある程度学説的に固まった内容を記述することとした。また、時代の流れに追従できるよう、分野ごとの巻の割り当てを見直した。旧音響工学講座では、共通する基礎の部分を除くと、6つの分野、すなわち電気音響、建築音

ii 「音響学講座」発刊にあたって

響，騒音・振動，聴覚と音響心理，音声，超音波から成り立っていたが，そのうち，当時社会問題にもなっていた騒音・振動に2つの巻を割いていた。本講座では，昨今の日本音響学会における研究発表件数などを考慮し，騒音・振動に関する記述を1つの巻にまとめる代わりに，音声に2つの巻を割り当てた。さらに，音響工学講座では扱っていなかった音楽音響を新たに追加すると共に，これからの展開が期待される分野をまとめた第10巻「音響学の展開」を刊行することとし，新しい技術の紹介にも心がけた。

本講座のような音響学を網羅・俯瞰する書籍は，国際的に見ても希有のものと思われる。本講座が，音響学を学ぶ諸氏の一助となり，また音響学の発展にいささかなりとも貢献できることを，心から願う次第である。

2019年1月

安藤彰男

「音響学講座」の全体構成は以下のようになっている。

- 第1巻 基礎音響学
- 第2巻 電気音響
- 第3巻 建築音響
- 第4巻 騒音・振動
- 第5巻 聴覚
- 第6巻 音声（上）
- 第7巻 音声（下）
- 第8巻 超音波
- 第9巻 音楽音響
- 第10巻 音響学の展開

ま え が き

近年、インターネットの普及に伴い、ネット上でのコミュニケーション社会が拡大している。とはいえ、やはり人間にとって最も自然なコミュニケーション手段は音声であり、家族や友人、同僚など周囲の人との日常のコミュニケーションの大半を、音声を介して行っているだろう。これまで産業界においても、音声をを用いた端末入力サービスなどが多く発表されてきており、音声研究の成果が世の中で活用されている。ただし、その利用環境には制限(条件)があり、まだまださまざまな課題が残されており、今後の音声研究への期待は大きいだろう。

さて本書では、音声の教科書として利用できるように、前半では、基礎的な音声分析アルゴリズム、音声生成メカニズムなどについても詳細に述べている。また後半では、応用研究として音声合成、雑音除去について説明している。

本書は六つの章からなる。1章では音声研究の分野への導入部であり、歴史と現状なども含めて簡潔に述べている。2章では音声信号処理の基礎について、音の高さや音声中に含まれる音韻情報などについて説明している。3章では人の発声の仕組みについて音声生成の音響メカニズムなど物理的な側面から説明している。またMRIで撮像した頭部断層像も用いて詳細に説明している。4章では音声波形の性質に注目した基礎的な音声分析法について説明している。5章では人と機械(ロボット)との情報交換手段として必要となる音声合成について述べている。近年の学習理論および計算機スペックの高性能化に伴い、機械から生成される音声品質は格段と自然性が改善されている。音声処理のみならず言語処理、および音声合成の応用研究、深層学習に基づく音声合成についても述べられている。6章では実際の環境で音声認識や音声通話を行う際に少なからず問題となる背景雑音、および残響の対処方法について説明している。ここでは特に単一マイクロホンでの処理に焦点を当てており、複数のマイクロホンをを用いたマイクロホンアレイ処理については、本シリーズ音響学講座第2巻

『電気音響』を参照されたい。また、音声知覚などについては、第5巻『聴覚』も参照されたい。音声認識、音声対話に関しては音響学講座の第7巻『音声（下）』で述べている。

最後に、本書が音声の分析・解析アルゴリズムの理解に役立ち、これからの音声処理の研究発展に少しでも貢献できれば望外の喜びである。

執筆分担は以下のとおりである。

- 有木康雄 1章, 2章, 4章
- 鏑木時彦 3章
- 戸田智基 5.1節~5.3節
- 南角吉彦 5.4節, 5.5節
- 藤本雅清 6.1節, 付録
- 木下慶介 6.2節

2021年7月

滝口哲也

目 次

1章 音声と音声研究の歴史

1.1 音 声 と は	2
1.1.1 人が音声を生成する過程	2
1.1.2 音声に対する雑音とひずみ	3
1.1.3 人が音声を認識する過程と冗長性	4
1.2 音声研究の歴史と現状	5
1.2.1 機械による音声の認識と生成	5
1.2.2 音声の分析合成の歴史と現状	8
1.2.3 音声合成の歴史と現状	9
1.2.4 音声認識の歴史と現状	11
1.3 音声研究と関連分野	17
1.3.1 音声研究と関連する分野	17
1.3.2 音声研究の応用分野	18
引用・参考文献	21

2章 音声波形の特性

2.1 音声の波形とスペクトル	25
2.1.1 音声波形の特徴	25
2.1.2 音声スペクトルの特徴	27
2.1.3 音声生成モデル	30
2.2 音韻の種類と特徴	31
2.2.1 音 韻 の 種 類	31
2.2.2 母音の特徴と分類	33

2.2.3 子音の特徴と分類	34
引用・参考文献	37

3章 音声の生成機構とそのモデル

3.1 音源-フィルタ理論：音声生成の基本モデル	39
3.1.1 音声コミュニケーションの成り立ち	39
3.1.2 音源とフィルタ：音声生成の基本要素	39
3.1.3 音声生成過程の線形モデル	41
3.1.4 音源-フィルタ理論と音源-フィルタ相互作用	43
3.2 声道の働きと音響モデル	45
3.2.1 人の音声器官	45
3.2.2 発話における声道の働き	46
3.2.3 声道の音響モデル	49
3.2.4 音響感度関数	54
3.2.5 声道形状の個人差	56
3.3 声帯音源のメカニズム	58
3.3.1 声帯音源の特徴	58
3.3.2 声帯音源の生成：発声の Myoelastic-Aerodynamic 理論	60
3.3.3 基本周波数の調整と発声時の声帯振動の様子	63
3.3.4 音源-フィルタ相互作用が発声に及ぼす影響	65
3.3.5 声帯音源の物理モデル	68
3.4 音声生成の生理学的過程	72
引用・参考文献	73

4章 音声の分析

4.1 スペクトル分析	76
4.1.1 スペクトル分析の目的	76
4.1.2 スペクトル分析の方法	77

4.1.3	ノンパラメトリックなスペクトル分析	78
4.2	ケプストラム分析	86
4.2.1	準同型分析とケプストラム分析	86
4.2.2	ケプストラム分析	90
4.2.3	複素ケプストラム分析	92
4.2.4	実ケプストラムと複素ケプストラムの関係	95
4.2.5	TANDEM-STRAIGHT	101
4.3	線形予測分析	105
4.3.1	線形予測分析の時間領域での定式化	107
4.3.2	予測係数の高速推定法	119
4.3.3	線形予測分析の最尤推定法による定式化	120
4.4	PARCOR 分析	130
4.4.1	PARCOR 分析の原理	130
4.4.2	PARCOR 分析の定式化	131
4.4.3	PARCOR 分析・合成システムの構成	141
4.4.4	PARCOR 係数と反射係数の関係	146
4.5	LSP 分析	151
4.5.1	LSP 分析の原理	151
4.5.2	LSP 分析の定式化	152
4.5.3	LSP 音声合成システムの構成	158
	引用・参考文献	162

5章 音声合成

5.1	音声合成の概要	167
5.1.1	音声合成の発展の歴史	167
5.1.2	人工的に音声を合成するには？	170
5.2	テキスト音声合成の基本的な仕組み	173
5.2.1	基本的な構成	173

5.2.2	言語処理部	174
5.2.3	音声処理	178
5.3	波形接続に基づく音声合成	183
5.3.1	素片選択アルゴリズム	184
5.3.2	コスト関数設計	189
5.3.3	音声コーパス設計	193
5.3.4	波形接続に基づく音声合成の利点と欠点	196
5.4	隠れマルコフモデルに基づく音声合成	196
5.4.1	隠れマルコフモデル	199
5.4.2	音声合成のための要素技術	202
5.4.3	柔軟な音声合成	213
5.4.4	音質改善手法	222
5.5	ニューラルネットワークに基づく音声合成	231
5.5.1	ニューラルネットワーク	231
5.5.2	深層音響モデルに基づく音声合成	235
5.5.3	深層波形生成モデルに基づく音声合成	241
5.5.4	End-to-End 型音声合成	243
	引用・参考文献	245

6章 音声処理の雑音対策

6.1	背景雑音の除去	258
6.1.1	雑音環境下での収音	258
6.1.2	ウィナーフィルタ	260
6.1.3	スペクトルサブトラクション	266
6.1.4	MMSE-STSA	273
6.1.5	信号部分空間法	278
6.1.6	より発展的な雑音除去法	281
6.2	残響の除去	283
6.2.1	残響とは	283

6.2.2 残響除去とは	285
6.2.3 種々の残響除去方法	286
引用・参考文献	292
付 録	298
A.1 直交原理の証明	298
A.1.1 最小二乗推定と擬似逆行列	298
A.1.2 直交射影	299
A.1.3 直交原理	300
A.2 複素関数の微分	301
A.2.1 微分の定義	301
A.2.2 複素関数微分への拡張	302
引用・参考文献	304
索 引	305

〈音声（下）目次〉

1章 音声認識

- 1.1 音声認識の概要
- 1.2 音声認識の原理
- 1.3 音声認識のための特徴量の抽出
- 1.4 音声認識システムの構成
- 1.5 音声認識結果の扱い

2章 音響モデルとその高度化

- 2.1 音響モデル
- 2.2 頑健性の向上
- 2.3 識別学習の利用
- 2.4 深層学習の利用

3章 言語モデルとその高度化

- 3.1 言語モデル
- 3.2 N -gram 言語モデル
- 3.3 統計的言語モデルの評価
- 3.4 頑健性の向上
- 3.5 識別的言語モデル
- 3.6 ニューラルネットワーク言語モデル

4章 話者認識

- 4.1 話者認識の概要
- 4.2 話者認識技術の進展と位置づけ
- 4.3 話者性の表現
- 4.4 話者認識システムの評価
- 4.5 話者ダイアライゼーション

5章 音声対話システム

- 5.1 対話システムのバリエーション
- 5.2 対話の主導権
- 5.3 対話管理のモデル
- 5.4 対話戦略の学習
- 5.5 音声対話システムの評価

1 章

音声と音声研究の歴史

◆本章のテーマ

この章では、人が日々話している音声には、多くの変形が生じていることを、音声生成の立場から考えてみる。つぎに、変形した音声を、人はどのような知識を用いて認識しているか考察する。これを基に、コンピュータによる音声認識、音声合成、分析合成の概要を学び、コンピュータの出現以降、音声の研究分野が発展してきた歴史を概観する。携帯電話で用いられている分析合成の技術は、音声認識と音声合成の発展に大きく寄与してきたことを学ぶ。音声合成では規則による音声合成、コーパスベース音声合成、統計的音声合成、ニューラル音声合成へと発展してきた歴史を概観する。音声認識では、単語音声認識から大語彙連続音声認識に至るまでの歴史と、その中で生じたブレイクスルーと深層学習について概観する。最後に、音声研究を支えている研究分野と、音声研究によってもたらされる応用分野について理解を深める。

◆本章の構成（キーワード）

1.1 音声とは

音韻情報、話者情報、感性情報、音声生成過程、雑音とひずみ、音声認識過程、冗長性

1.2 音声研究の歴史と現状

音声認識、音声合成、分析合成、音声検索、音声翻訳、音響モデル、言語モデル、声質変換、音声対話、ボコーダ、最尤スペクトル推定法、規則による合成、コーパスベース音声合成、HMM 音声合成、ニューラル音声合成、動的計画法、2 段 DP、連続音声認識、隠れマルコフモデル、トライグラム、大語彙連続音声認識、音声理解システム、自由発話の音声認識、音声ドキュメント検索、音声翻訳システム、音声対話システム、End-to-End

1.3 音声研究と関連分野

信号処理、音響学、音声学、言語学、音韻論、情報科学、生理学、心理学、音声変換、オフラインテキスト変換、リアルタイム認識合成

1.1 音声とは

1.1.1 人が音声を生成する過程

われわれの社会生活では、自分の考えを述べたり、人に依頼をしたり、知りたいことを尋ねたり、雑談・経験を語って共感するといった言語活動が頻繁に発生する。こういった言語活動は、古くは手紙、最近ではメールやチャットといった文字を用いて行われることもあるが、通常、対面している人に対しては音声を用いて行われる。したがって、音声は人にとって最も自然で、容易かつ効率的で、一般に広く用いられている伝達手段であるといえる。

このような音声には、どのような情報が含まれているのであろうか。電話がかかってきて話をしている場合を想定してみよう。誰からの電話なのか、どのような内容で、先方はどんな感情をもっているのか、われわれはこれらの情報を容易に聞き分けることができる。音声にはこのように、誰が話をしているかといった話者情報、なにを話しているかといった音韻情報、どのような感情で話をしているかといった感性情報の三つの情報が含まれている。

人はこのような音声を、どのような過程を経て生成し、認識しているのであろうか。図 1.1 の下半分は、音声生成の過程を概念的に表したものである。意味や概念を形成する空間で、考えや意図 M が生じると、それは単語の系列であるテキスト W として生成される。つぎに、そのテキストは、音素 (phoneme) とその継続時間長、アクセントや抑揚の系列に変換され、さらにそれは、調音器官 (articulatory organs) を動かして発音させるために、脳からの指令の系列に変換される。この指令に従って、調音器官が連続的に運動し、連続的な空

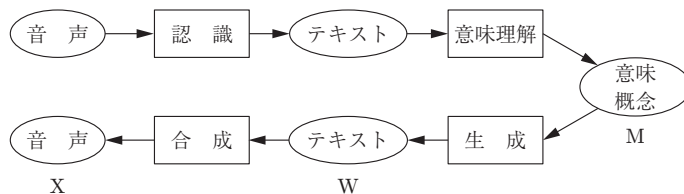


図 1.1 人が音声を認識し生成する過程

気の振動が音声 X として生成される^{1)†}。

1.1.2 音声に対する雑音とひずみ

1.1.1 項に述べたことは、言語と関係する音韻情報を生成する過程であるが、調音器官は個人によってその形状と動きが異なるため、音韻情報の上に話者情報が付加される。さらに、伝えようとする考えや意思に付随して、喜びや怒りといった感性情報も付加される。こうして多様な情報をもつ音声生成されることになるが、基になる音韻情報を忠実に伝送するという通信の立場から見ると、話者情報や感性情報は雑音・ひずみとみなすこともできる。

音韻情報に付加される雑音・ひずみは、話者情報や感性情報だけではない。音声生成の過程は、調音器官の連続的な筋肉運動により、離散的な音韻系列を連続的な信号に変換する過程である。したがって、いくつかの音素がつづけて発音されるとき、調音器官の運動速度が有限であることから、調音器官が一つの音素の目標状態に至る前に、つぎの音素に移るいわゆる「なまけ」の現象が生じる。これを調音結合 (co-articulation) というが、これにより連続音声中の音素は大きな変形を受けることになる。

また、発話の仕方によっても音声は変形を受ける。例えば、ニュース音声のような原稿を読み上げたような音声は、明瞭であり聞きやすい。しかし、インタビューや講演などのように原稿を読み上げるのではなく、考えながら話をする自然発話 (spontaneous speech) の音声は、聞き取ることはできるがそれほど明瞭ではない。さらに、日常的な会話音声となると、情緒性が強くなり明瞭性は低下する。このような発声においては、言語的に、倒置、繰返し、不要語の付加、未知語、言い間違い、言いよどみといったひずみが発生する。

さらに、発声した音声は空気中を伝搬して聞き手に伝わるので、その間に発生する雑音や人の声、残響などの影響を受けてひずむ。マイクやスピーカを使用している場合には、それらの機器や伝送路によってひずみを受ける。このよ

† 肩付き数字は、章末の引用・参考文献の番号を表す。

うに、音声が生産されて伝達される過程には、多くの雑音・ひずみが含まれている。

1.1.3 人が音声を認識する過程と冗長性

人は、1.1.2項で述べたような種々の雑音やひずみによって変形された音声を取り取り、話者の考えや意図を把握している。図 1.1 の上半分に示した音声の認識では、音韻、単語を認識し、さらに文節、文を認識し、一連の文のつながりから文脈や意味を理解し、現実の事物と対応づけて話者の考えや意図を理解していると思われる。これは人が音声認識している過程であるが、一体どうやって、雑音やひずみによって変形した音声から話者の考えや意図を復元しているのかという疑問がわいてくる。

一つの考え方は、冗長性である。無限の集合があり、その要素の組合せに制約がなく、すべての組合せが意味をもつ場合には冗長性はない。要素の連鎖の一部が変形すれば意味が変わるからである。しかし、音素も単語も有限の集合であり、組合せには制約がある。その連鎖の一部が変形しても、有限の要素のどれかにマッピングされ、組合せの制約によって、変形前の要素に復元可能であろう。

しかも、現実的に意味をもつ文となると、有限の要素で制約のある組合せは、さらに限定される。冗長性に付随するこのような制約は、辞書や文法といった言語的知識、現実世界の意味的知識として、明示的に記述することができる。われわれは、意識的・無意識的にこのような知識を用いて言語を認識していると考えられる。

音声信号は連続信号であり、一部に雑音が重畳したり、変形したりしても聞き取ることが可能である。単語音声の子音部分を雑音に置き換えても、人は単語を認識できるといわれている²⁾。このように、音声や言語には冗長性が存在している。人は、雑音・ひずみを伴って生成された音声を取り、音声や言語に内在している冗長性を利用して、話者の考えや意図を把握しているものと考えられる¹⁾。

1.2 音声研究の歴史と現状

1.2.1 機械による音声の認識と生成

〔1〕 音声インタフェース コンピュータの出現とともに情報化社会が進展し、ネットワークやデータベース、クラウドコンピューティングが整い、最近では、IoT や人工知能が展開する社会となってきた。このような状況で、人とコンピュータはより密接な関係となり、人のよきパートナーとして、高いコミュニケーション能力がコンピュータに求められている。

すでに述べたように、音声は人にとって最も自然で、容易かつ効率的な伝達手段であり、広く一般に用いられている。したがって、音声でコンピュータとコミュニケーションできることは、人にとってごく普通の要求である。こうして、コンピュータが出現して以来、文字中心のテキストインタフェースに加えて、マウスやアイコンといったグラフィカルインタフェースが出現し、最近では、音声や顔画像といった人の五感に関わるマルチメディアインタフェースが用いられるようになってきた³⁾。

また、コンピュータのダウンサイジングにより、携帯端末や携帯電話のようなモバイルコンピュータ、さらに眼鏡や時計型のウェアラブルコンピュータが出現し、AI スピーカや家電製品のような「コンピュータの形をとらないコンピュータ」も浸透している。いわゆるユビキタスコンピューティングの時代であり、人と接するこのような多くの製品に、音声インタフェースが搭載されている。

〔2〕 音声研究の分野 携帯電話は、最も優れた音声インタフェースを有する装置であろう。それは、音声通信、音声認識 (speech recognition)、音声合成 (speech synthesis) を基本的にもっており、アプリケーションによって音声検索 (voice search) や音声翻訳 (speech translation) も可能である。音声認識とは、人が発した言語音声を対応するテキストに変換する技術であり、音声合成はその逆で、任意のテキストから対応する音声を人工的に生成する技術のことである。

索引

【あ】

アクセント核	176
アクセント型	176
安定性	145

【い】

異音	32
位相ベクトル	81, 267
位相復元	95
一般化スペクトルサブ トラクション	270
入りわたり	36
因果性	94, 96
因果性数列	99
因果的インパルス応答	284
因子分析	216
咽頭腔	45
インパルス応答	26

【う】

ウィーナー・ヒンチンの 定理	262
ウィーナーフィルタ	260
ウェブスター方程式	49
後向き予測	131
後向き予測係数	131

【え】

エルゴード性	264
エンコーダ・デコーダ ネットワーク	245

【お】

オーバーサブトラクション 係数	267
音韻情報	2
音響特徴	6
音響モデル	6

音源信号	8
音源-フィルタ理論	40
音声器官	46
音声記号	32
音声区間検出	270
音声検索	5
音声合成	5
音声生成の過程	2
音声生成モデル	30
音声対話	7, 16
音声ドキュメント検索	16
音声認識	5
音声波形	25
音声表記	32
音声翻訳	5
音声翻訳システム	16
音声理解	13
音素	32
音素継続長	178
音素表記	32

【か】

開大期	59
開放期	59
下顎	45
過学習	235
確率的勾配降下法	233
確率モデルに基づく統計的 パラメトリック音声合成	169
隠れセミマルコフモデル	207
隠れマルコフモデル	11, 13, 196, 199, 282
活性化関数	232
カルマンフィルタ	270
感情認識	7
感性情報	2

【き】

機械学習	282
機械式音声合成器	167
規則による合成	10
基本周期	26
基本周波数	26, 28
基本周波数パターン	178
逆Z変換	93
逆短時間フーリエ変換	82
逆特性	286
逆フィルタ	109, 133
逆離散時間フーリエ変換	88
逆離散フーリエ変換	82, 264
共分散行列	113
共分散法	112
共鳴	41
極	78
極零モデル	107

【く】

偶関数	96
空間相関行列	287
唇	45
クラスタ適応学習	216
クロススペクトル	262
クロスリンガル話者適応	219

【け】

形態素	175
系列内変動モデリング	222
ケプストラム	88
ケプストラム分析	78, 88
ケプストラム法	77
言語特徴	6
言語モデル	6

【こ】			
口腔	45	残響除去	285
硬口蓋	45	残差ブロック	242
格子形フィルタ	142	【し】	
高次ケフレンシー	91	子音	26
後進波	147	自己回帰-移動平均モデル	107
合成単位	171	事後確率最大化基準	226, 278
合成フィルタ	109	自己相関関数	110, 262
喉頭	46	自己相関行列	111
後部残響	284	自己相関法	111
呼吸器官	46	自然発話	3
誤差逆伝搬法	235	実ケプストラム	90
コーパスベース音声合成	10, 169	実時間リカレント学習	238
固有声	215	室内伝達関数モデル	286
固有値分解	278	重回帰 HMM	216
混合過程	258	自由発話の音声認識	15
混合正規分布モデル	14, 282	主成分分析	281
コンテキスト依存モデル	210	出力確率	199
コンテキスト加算モデル	226	受容野	241
コンテキストクラスタリング	211	準同型分析	87
——に基づく手法	181	準同型変換	87
【さ】		準同型法	78
最小位相数列	94, 98	準同型ボコーダ	101
最小二乗推定法	109	条件異音	32
最適推定値	260	条件付き確率密度関数	274
最適フィルタ	260	状態遷移確率	199
サイドローブ	80	状態遷移トレリス	200
最尤推定法	120	冗長性	4
最尤スペクトル距離	130	初期状態確率	199
最尤スペクトル推定法	9	初期反射音	284
最尤線形回帰	213	自励振動	40
最尤ボコーダ	121	事例ベース音声強調法	282
サウンドスペクトログラム	28	信号対雑音比	267
雑音除去	258	信号復元過程	260
雑音重畳過程	260	信号部分空間法	279
残響	283	深層学習	16, 282
		振幅スペクトル	28, 81, 266
		【す】	
		推定オピニオン評点	194
		スパース性	289
		スペクトル	8
		スペクトルサブトラク ション	265, 266
		スペクトル包絡	28
		スペクトル補償	105
		スペクトル密度関数	120
		スペクトログラムリー ディング	37
		【せ】	
		正規化角周波数	82, 88
		正規方程式	110
		声質変換	7
		声帯	28, 40, 59
		声帯音源	40
		声帯振動	26
		声帯振動音源	26
		静的特徴量	203
		声道	29, 41, 45
		声道フィルタ	43, 109
		声門	29, 40, 59
		舌	45
		全極モデル	78
		線形予測符号化法	9
		線形予測分析	78, 106, 109
		前進波	147
		線スペクトル対分析	9, 78
		【そ】	
		相互相関関数	262
		双方向 RNN	238
		ソース・フィルタモデル	167
		素片選択	169
		損失関数	233, 273
		【た】	
		帯域幅	116
		大語彙音声認識	14
		対称テブリッツ行列	111
		多管近似	50
		多空間上の確率分布に基 づく HMM	209
		畳み込み	26

畳み込みニューラル
ネットワーク 241

単音 32

短時間スペクトル 81

短時間フーリエ変換 81, 290

【ち】

チャンネルボコーダ 85

中心周波数 116

調音位置 35

調音器官 29, 45

調音結合 3

調音様式 35

調波成分 28

直接音 283

直交原理 135, 261

【て】

定在波 41

低次ケフレンシー 91

定常過程 260

ディープニューラル
ネットワーク 231

適応 213

適合誤差 128

テキスト音声合成 168

出わたり 36

電気式音声合成器 168

電話 167

【と】

統計の音声合成 11

統計的機械翻訳 16

統計的言語モデル 14

統計モデル 282

同時確率密度関数 274

同時事後確率最大化基準 278

動的計画法 13

動的特徴量 203

特異値分解 279

特徴量空間 278

独立成分分析 281

トライフォンモデル 210

トラジェクトリ HMM 225

【な】

軟口蓋 45

【に】

2段 DP 13

ニューラルネットワーク 282

——に基づく音声合成 169

【は】

背景雑音 258

拍 176

白色性 279

波形接続型音声合成 169

バックプロパゲーション 235

発語 39

発声 39

——の Myoelastic-
Aerodynamic 理論 60

発声器官 46

発話 39

パーティクルフィルタ 270

波動方程式 147

パラメトリック法 78

破裂音源 40

パワーケプストラム 90

パワースペクトル 81, 262

パワーパターン 178

反射係数 149

半正定値行列 111

判定帰還 265

バンドパスフィルタ 77

【ひ】

鼻腔 45

非線形スペクトルサブト
ラクション 271

非定常過程 281

非負値行列因子分解 282

非負値行列逆畳み込み 288

ピリオドグラム 128

【ふ】

フィルタ内積 133

フィルタバンク 77, 82

フォルマント 28, 41

フォルマント音声合成 168

フォルマント周波数 28

フォルマントボコーダ 151

負帰還路 158

複素ガウス分布 291

複素ケプストラム 92

複素数スペクトル 263

複素正規分布 275

符号帳 106, 282

符号励振線形予測符号化 106

藤崎モデル 180

部分観測マルコフ決定
過程 16

部分空間法 278

部分自己相関係数 131

ブライント残響除去 285

フレーム 79

フレーム周期/シフト 79

フレーム長 79

フロアリング係数 267

分析合成 7

分析フィルタ 109, 133

【へ】

平均オビニオン評点 190

平均二乗誤差 260

閉鎖期 59

閉小期 59

ベイズ基準 226

ベイズ推定 273

ベルヌーイの定理 62

偏自己相関係数 131, 134

偏自己相関分析 9, 78

弁別素性 32

【ほ】

母音 26

ボコーダ 8, 101, 167

ポリグロット音声合成	219	メル尺度	85	離散時間フーリエ変換	87
【ま】		メルフィルタバンク	85	離散フーリエ変換	81, 262
前向き予測	131	【も】		粒子フィルタ	270
前向き予測係数	131	モーラ	176	量子化特性	106
摩擦音源	40	【ゆ】		【る】	
窓関数	79, 263	有色性	281	ルールベース音声合成方式	168
マルコフ連鎖モンテカルロ	270	有声音	30	【れ】	
【み】		尤度最大化基準	226	励振源モデル	227
ミニバッチ SGD 法	233	ユール・ウォーカーの方		連結学習	202
ミュージカルノイズ	271	程式	110	連続音声認識	14
【む】		【よ】		【わ】	
無声音	31	予測係数	108	話者情報	2
【め】		予測誤差	108	話者適応学習	214
メインローブ	80	予測回数	109	話者認識	7
メルケプストラム係数	11	【り】			
		リカレントニューラル			
		ネットワーク	237		
◇					
【A】		DP	13	GMM-HMM	14
AR-MA モデル	107	DTFT	87	GRU	240
【B】		【E】		GV モデリング	222
Baum-Welch アルゴリズム	201	EM アルゴリズム	202	【H】	
BPTT	238	End-to-End 型音声合成	170	HMM	11, 196, 282
【C】		End-to-End モデル	16	HMM 音声合成	11, 197
CAT	216	【F】		HSMM	207
CELP 符号化	106	FA	216	【I】	
CNN	241	Feed-Forward ニューラルネットワーク	231	IDFT	82, 264
COC 法	181	FFNN	231	IDTFT	88
Code Book	282	Forward-Backward アルゴリズム	201	ISTFT	82
【D】		【G】		Itakura-Saito 距離	130
DD	265	Generalized-Forward-Backward アルゴリズム	207	【J】	
DFT	81, 262	GMM	14, 282	JMAP 基準	278
Dilated Convolution	242			【K】	
DNN	231			KL ダイバージェンス	215

【L】		MOS	190	【S】	
		MSD-HMM	209		
Left-to-Right 型	199	MVDR ビームフォーマ	288	SAT	214
Levinson-Durbin 法	119	【N】		SDR	16
LPC	9			SGD 法	233
LPC 分析	78	NMF	282	SMT	16
LPC ボコーダ	119	NMFD	288	SNR	267
LSP 音声合成フィルタ	160	【P】		SPSS	11
LSP 音声分析合成方式	151			STFT	81, 290
LSP 分析	9, 78	PARCOR 係数	131	STRAIGHT	101
LSTM	239	PARCOR 合成システム	143	SVD	279
【M】		PARCOR 分析	9, 78	【T】	
MAP 基準	226, 278	PARCOR 分析システム	142	TANDEM-STRAIGHT	
MFCC	11	PARCOR ボコーダ	144		101
Minimum Statistics	270	POMDP	16	【W】	
Mixture Density Network		【R】			
	236	Receptive Field	241	WaveNet	241
ML 基準	226	RNN	237	【Z】	
MLLR	213	RTRL	238		
MMSE-STSA	273			Z 変換	93

——編著者・著者略歴——

滝口 哲也（たきぐち てつや）

- 1994年 岡山理科大学理学部応用数学科卒業
- 1996年 奈良先端科学技術大学院大学情報科学研究科博士前期課程修了（情報処理学専攻）
- 1999年 奈良先端科学技術大学院大学情報科学研究科博士後期課程修了（情報処理学専攻）
博士（工学）
- 1999年 IBM 東京基礎研究所副主任研究員
- 2004年 神戸大学講師
- 2009年 神戸大学准教授
- 2017年 神戸大学教授
- 現在に至る

有木 康雄（ありき やすお）

- 1974年 京都大学工学部情報工学科卒業
- 1976年 京都大学大学院工学研究科修士課程修了（情報工学専攻）
- 1979年 京都大学大学院工学研究科博士課程単位取得退学（情報工学専攻）
工学博士
- 1990年 龍谷大学助教授
- 1992年 龍谷大学教授
- 2003年 神戸大学教授
- 2016年 神戸大学名誉教授

鍋木 時彦（かぶらぎ ときひこ）

- 1984年 九州芸術工科大学芸術工学部音響設計学科卒業
- 1986年 九州芸術工科大学大学院芸術工学研究科修士課程修了（情報伝達専攻）
- 1986年 日本電信電話株式会社基礎研究所勤務
- 1998年 九州芸術工科大学大学院芸術工学研究科博士課程修了（情報伝達専攻）
博士（芸術工学）
- 2000年 九州芸術工科大学助教授
- 2003年 九州大学大学院助教授
- 2007年 九州大学大学院准教授
- 2015年 九州大学大学院教授
- 現在に至る

戸田 智基（とだ ともき）

- 1999年 名古屋大学工学部電気電子・情報工学科卒業
- 2001年 奈良先端科学技術大学院大学情報科学研究科博士前期課程修了（情報処理学専攻）
- 2003年 奈良先端科学技術大学院大学情報科学研究科博士後期課程修了（情報処理学専攻）
博士（工学）
- 2003年 日本学術振興会特別研究員-PD
- 2005年 奈良先端科学技術大学院大学助手
- 2007年 奈良先端科学技術大学院大学助教
- 2011年 奈良先端科学技術大学院大学准教授
- 2015年 名古屋大学教授
- 現在に至る

南角 吉彦 (なんかく よしひこ)

- 1999年 名古屋工業大学工学部知能情報システム学科卒業
- 2001年 名古屋工業大学大学院工学研究科博士前期課程修了 (電気情報工学専攻)
- 2004年 名古屋工業大学大学院工学研究科博士後期課程修了 (電気情報工学専攻) 博士 (工学)
- 2004年 名古屋工業大学テクノイノベーションセンター 研究員
- 2005年 名古屋工業大学 助手
- 2007年 名古屋工業大学 助教
- 2012年 名古屋工業大学 准教授
- 現在に至る

藤本 雅清 (ふじもと まさきよ)

- 1997年 龍谷大学理工学部電子情報学科卒業
- 2001年 龍谷大学大学院理工学研究科修士課程修了 (電子情報学専攻)
- 2004年 龍谷大学大学院理工学研究科博士課程単位取得退学 (電子情報学専攻)
- 2004年 株式会社国際電気通信基礎技術研究所音声言語コミュニケーション研究所勤務
- 2005年 博士 (工学) (龍谷大学)
- 2006年 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所勤務
- 2016年 情報通信研究機構主任研究員
- 現在に至る

木下 慶介 (きのした けいすけ)

- 2001年 上智大学理工学部電気電子工学科卒業
- 2003年 上智大学大学院理工学研究科修士課程修了 (電気電子工学専攻)
- 2004年 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所勤務
- 現在に至る
- 2010年 博士 (工学) (上智大学)

音声 (上)

Speech Vol.1

© 一般社団法人 日本音響学会 2021

2021年9月22日 初版第1刷発行

検印省略

編者 一般社団法人 日本音響学会
発行者 株式会社 コロナ社
代表者 牛来真也
印刷所 三美印刷株式会社
製本所 有限会社 愛千製本所

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

ISBN 978-4-339-01366-5 C3355 Printed in Japan

(金)



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いません。
落丁・乱丁はお取替えいたします。