

メディア学大系

13

音声音響インタ フェース実践

相川 清明

大淵 康成

共著



コロナ社

メディア学大系 編集委員会

監修（第1期）

相川 清明（東京工科大学，工学博士）
飯田 仁（東京工科大学，博士（工学））

監修（第2期）

相川 清明（東京工科大学，工学博士）
近藤 邦雄（東京工科大学，工学博士）

編集委員

稲葉 竹俊（東京工科大学）
榎本 美香（東京工科大学，博士（学術））
太田 高志（東京工科大学，博士（工学））
大山 昌彦（東京工科大学）
柿本 正憲（東京工科大学，博士（情報理工学））
菊池 司（東京工科大学，博士（工学））
近藤 邦雄（東京工科大学，工学博士）
榎 俊吾（東京工科大学，博士（社会情報学））
佐々木 和郎（東京工科大学）
進藤 美希（東京工科大学，博士（経営管理））
寺澤 卓也（東京工科大学，博士（工学））
三上 浩司（東京工科大学，博士（政策・メディア））

「メディア学大系」刊行に寄せて

ラテン語の“メディア（中間・仲立ち）”という言葉は、16世紀後期の社会で使われ始め、20世紀前期には人間のコミュニケーションを助ける新聞・雑誌・ラジオ・テレビが代表する“マスメディア”を意味するようになった。また、20世紀後期の情報通信技術の著しい発展によってメディアは社会変革の原動力に不可欠な存在までに押し上げられた。著名なメディア論者マーシャル・マクルーハンは彼の著書『メディア論——人間の拡張の諸相』（栗原・河本 訳、みすず書房、1987年）のなかで、“メディアは人間の外部環境のすべてで、人間拡張の技術であり、われわれのすみからすみまで変えてしまう。人類の歴史はメディアの交替の歴史ともいえ、メディアの作用に関する知識なしには、社会と文化の変動を理解することはできない”と示唆している。

このように未来社会におけるメディアの発展とその重要な役割は多くの学者が指摘するところであるが、大学教育の対象としての「メディア学」の体系化は進んでいない。東京工科大学は理工系の大学であるが、その特色を活かしてメディア学的一端を学部レベルで教育・研究する学部を創設することを検討し、1999年4月に先駆けて「メディア学部」を開設した。ここでいう、メディアとは「人間の意思や感情の創出・表現・認識・知覚・理解・記憶・伝達・利用といった人間の知的コミュニケーションの基本的な機能を支援し、助長する媒体あるいは手段」と広義にとらえている。このような多様かつ進化する高度な学術対象を取り扱うためには、従来の個別学問だけで対応することは困難で、諸学問横断的なアプローチが必須と考え、学部内に専門的な科目群（コア）を設けた。その一つ目はメディアの高度な機能と未来のメディアを開拓するための工学的な領域「メディア技術コア」、二つ目は意思・感情の豊かな表現力と秘められた発想力の発掘を目指す芸術学的な領域「メディア表現コ

ii 「メディア学大系」刊行に寄せて

ア], 三つ目は新しい社会メディアシステムの開発ならびに健全で快適な社会の創造に寄与する人文社会学的な領域「メディア環境コア」である。

「文・理・芸」融合のメディア学部は創立から13年の間、メディア学の体系化に試行錯誤の連続であったが、その経験を通して、メディア学は21世紀の学術・産業・社会・生活のあらゆる面に計り知れない大きなインパクトを与え、学問分野でも重要な位置を占めることを知った。また、メディアに関する学術的な基礎を確立する見通しもつき、歴年の願いであった「メディア学大系」の教科書シリーズ全10巻を刊行することになった。

2016年に至り、メディア学の普及と進歩は目覚ましく、「メディア学大系」もさらに増強が必要になった。この度、視聴覚情報の新たな取り扱いの進歩に対応するため、さらに5巻を刊行することにした。また、学術・産業・社会の変革に貢献する斬新的なメディアに関する教科書を随時追加し、「メディア学大系」を充実させることを計画している。

この「メディア学大系」の教科書シリーズは、特にメディア技術・メディア芸術・メディア環境に興味をもつ学生には基礎的な教科書になり、メディアエキスパートを志す諸氏には本格的なメディア学への橋渡しの役割を果たすと確信している。この教科書シリーズを通して「メディア学」という新しい学問の台頭を感じとっていただければ幸いである。

2017年1月

東京工科大学
メディア学部 初代学部長
前学長

相磯秀夫

「メディア学大系」の使い方

メディア学は人から人や社会への情報伝達に関する学問である。したがって、メディア学は情報工学、文化、社会、人文科学にも及ぶ。「メディア学大系」の第1巻から第10巻においては、文系・理系の範ちゅうを超えたメディア学という新しい学問領域の全体像を学部学生に理解してもらうために、5領域から説明を行っている。

第1巻『メディア学入門』において、歴史的背景もふまえて、メディアの全体像とメディア学の学びの対象を概観している。

第2巻『CGとゲームの技術』、第3巻『コンテンツクリエイション』は、ゲーム、アニメなどコンピュータグラフィックスに関係した内容である。

第4巻『マルチモーダルインタラクション』、第5巻『人とコンピュータの関わり』は、人とコンピュータのコミュニケーションをインタフェースとそれに関係する技術の側面から記述している。

第6巻『教育メディア』、第7巻『コミュニティメディア』は、メディアの活用のうち、人と人のつながりや社会に関係した内容である。

第8巻『ICTビジネス』、第9巻『ミュージックメディア』は、メディアの活用のうち、サービスなどの産業、ビジネスや経済に関係した分野である。

第10巻『メディアICT』は、メディアを学ぶ基礎となるコンピュータの技術の入門書である。

今回追加した第11巻から第15巻は、人とコンピュータの間のメディア情報伝達の中心をなす視聴覚に重点を置いている。また、内容的には、第1巻から第10巻には書ききれなかった内容とプログラムなどを用いて、より实际的に理解する内容を含めた。

第11巻『自然現象のシミュレーションと可視化』では、高度なコンピュー

タグラフィック技術とメディアデータを視覚的に表現する方法について述べている。

第12巻『CG数理の基礎』は、コンピュータグラフィックスや画像処理の原理を、コンピュータツールやプログラムを用いて学ぶ。

第13巻『音声音響インタフェース実践』では、音と音声をコンピュータで扱う技術をコンピュータツールやプログラムを用いて学ぶ。

第14巻『映像メディアの制作技術』は、デジタル映像の制作と配信の技術を広告の領域まで含めて解説する。

第15巻『視聴覚メディア』は、アニメなどの映像や画像の処理と人の視覚特性、楽器音など自然界の音に対する聴覚特性など、人の視聴覚の側面から解説する。

各巻の構成内容は、半年にわたる15週、90分2単位の大学学部における授業を想定して執筆され、各章に演習問題を設置して自主学習の支援をするとともに、参考文献を適切に提示し、十分な理解ができるようにしている。

メディアの歴史は太古に溯るが、デジタルメディアの時代になり、新しい方法や技術がつつぎに導入され、急速な進展を続けている。本シリーズは、将来にわたって通用するメディア学の基本的な考え方の修得に重点を置いて企画した。また、メディア学はメディアの実体との結びつきが強いという特徴がある。そのため、各分冊の執筆にあたり、実践的な演習授業の経験が豊富で最新の展開を把握している第一線の執筆者を選び、執筆をお願いした。メディア学は限りなく進展する学問である。本書がメディアを志す読者の確固たる礎となることを期待する。

2017年1月

相川清明
近藤邦雄

まえがき

本書は「音声音響インタフェース実践」という一見演習の手引きのように思えるタイトルであるが、実践を通して基礎理論の理解を深めることを目的としている。「メディア学大系」第4巻「マルチモーダルインタラクション」の中で、音のデジタル信号処理の理論を解説しているが、本書は、そこには記載できなかった信号の基礎理論とその具体的応用、および信号処理を応用したインタフェースについて解説する。

2章では、特に具体的な数値を挙げて物理現象を実感できるように工夫した。複素数を含む数式や演算の意味をわかりやすく解説するようにしている。本書では、エコーキャンセラのような高度な処理についても触れているが、何をしたいのかから始まって、レベルの高い理論まで段階を追ってスムーズに導いている。ビームフォーマという特定の方向からの音を取り込む仕組みについては、難しい理論をわかりやすい図を多用して理解に結びつけている。ブライント音源分離や独立成分分析などの最先端の信号処理技術をも、その仕組みがわかりやすいように解説した。さらに、音場制御や騒音の除去の側面からも、基礎的な数式を用いながらも原理の理解に重点を置いた解説を行っている。

3章は、信号処理ツールを用いてデジタル信号処理を実感できる構成とした。東京工科大学のメディア学部はノートパソコン必携である。このため、授業でこのような信号処理ツールを活用できる。本書で用いている MATLAB や Scilab を用いると、簡単に音の入出力や生成加工ができる。さらに、なかなかイメージをつかみにくい複素関数の演算も簡単に行うことができる。これらのツールは描画能力にも優れているので、処理した結果を音だけでなく、さまざまな図に表現することができる。特にデジタルフィルタの演算や伝達関数の表示では、これらのツールは威力を発揮する。本章では、これらのツールを使

うことにより，理論を実感して理解できるようにした。

4章は，その他の音声音響処理用のツールと最先端の考え方の紹介である。特に，データから得られた特徴量の学習に基づく最近の音声認識や機械学習についてわかりやすく解説しており，最先端のディープラーニングにまで触れている。

本書は，1，2，4章を大淵が，3章を相川が担当した。これらの章は，段階を追って読まないで理解できないということはなく，どこから読み始めてもよい。1章は各章への導入部なので，まずそれを読んでいただき，必要に応じて各章に進んでいただくとよいと思われる。本書は実践を通して理論を理解するための書物であり，プログラムを記載している部分では，なるべく紹介しているツールを手元において，実践しながら読み進んでいただけると幸いである。

2017年1月

相川清明

大淵康成

目 次

1 章

音声音響インタフェースの実現のために

1.1 身の回りの音声音響インタフェース	2
1.2 ツールを活用したインタフェース実践	3
演習問題	5

2 章

音響インタフェース実現のための基礎知識

2.1 音の性質と周波数分析	7
2.1.1 音波の伝搬とエネルギー	7
2.1.2 音の振動と三角関数	9
2.1.3 波の重ね合わせとフーリエ変換	10
2.1.4 スペクトログラム	14
2.1.5 サンプリング	16
2.1.6 畳み込み演算と伝達関数	18
2.1.7 音の複素数表現	20
2.2 エコーキャンセラ	22
2.2.1 エコーの発生とハウリング	22
2.2.2 エコーサプレッサ	23
2.2.3 エコーキャンセラの原理	25
2.2.4 誤差最小化による解法	26
2.2.5 LMS アルゴリズム	28
2.2.6 非線形エコーキャンセラ	30
2.2.7 ダブルトーク検出	32
2.2.8 エコーキャンセラの実装	34

2.3	マイクロホンアレイ	35
2.3.1	複数のマイクで取り込んだ音の性質	35
2.3.2	適応ノイズキャンセラ	36
2.3.3	遅延和ビームフォーマ	38
2.3.4	死角形成型ビームフォーマ	43
2.3.5	適応ビームフォーマ	45
2.3.6	音源方向推定	48
2.3.7	非同期マイクロホンアレイ	51
2.4	ブラインド信号分離	52
2.4.1	周波数領域でのバイナリマスキング	52
2.4.2	独立成分分析	54
2.4.3	非負値行列因子分解	58
2.5	単一マイク信号からの雑音抑圧	60
2.5.1	スペクトルサブトラクション	60
2.5.2	統計的雑音抑圧	62
2.6	音場制御	63
2.6.1	インパルス応答と伝達関数の測定	63
2.6.2	ステレオ再生とサラウンド	65
2.6.3	バイノーラル録音	66
2.6.4	頭部伝達関数	68
2.6.5	アクティブノイズコントロール	70
2.6.6	スピーカアレイ	72
2.6.7	パラメトリックスピーカ	73
	演習問題	74

3章**MATLAB/Scilab による音声音響信号処理の実践**

3.1	音声音響信号の入出力と描画	77
3.1.1	本章におけるきまり	77
3.1.2	デジタル音信号の作成と出力	77

3.1.3	音の入力	80
3.1.4	ファイルへの保存と読み込み	81
3.1.5	波形とスペクトルの描画	82
3.2	デジタルフィルタ	84
3.2.1	通過帯域によるフィルタの分類	84
3.2.2	時間領域と周波数領域	86
3.2.3	フィルタ演算	89
3.2.4	FIRフィルタとIIRフィルタ	92
3.2.5	フィルタの周波数特性	93
3.2.6	極と零点	96
3.2.7	IIRデジタルフィルタ	101
3.2.8	収束する場合	102
3.2.9	IIRフィルタ出力が発散する場合	106
3.2.10	安定性	107
3.2.11	極が負の実数で重根の場合の伝達関数	112
3.2.12	極が複素数の場合	114
3.2.13	Q値が高いフィルタ	115
3.2.14	バターワースフィルタとチェビシェフフィルタ	118
3.3	効果音の生成	123
3.3.1	音の加工	123
3.3.2	ビブラート	123
3.3.3	倍音成分を含むビブラート	126
3.3.4	リバーブとエコー	128
3.4	スペクトル分析	131
3.4.1	スペクトログラム	131
3.4.2	窓関数	132
3.4.3	プリエンファシス	135
3.4.4	マトリックスの色表示	135
3.5	音声音響特有の信号処理	139
3.5.1	線形予測分析	139

3.5.2	ボコーダ	145
3.6	音声認識と音声合成のための基本演算	149
3.6.1	ケプストラム	149
3.6.2	ケプストラムによるピッチ抽出	150
3.6.3	変形相関関数によるピッチ抽出	151
3.6.4	音声認識における音響処理の基本	153
3.6.5	ユークリッド距離	154
3.6.6	cos 類似度	154
3.7	楽器音の合成	155
3.7.1	合成方式	155
3.7.2	VCO	157
3.7.3	VCA	157
3.7.4	ADSR	157
3.7.5	VCF	158
3.7.6	RG	159
3.7.7	LFO	159
3.7.8	シンセサイザのプログラム	160
	演習問題	163

4章

ツールキットを活用した音声音響信号処理と機械学習の実践

4.1	音響データ収集	166
4.1.1	音のデータを集める	166
4.1.2	声のバリエーション	168
4.1.3	音のデータを作る	169
4.1.4	公開データを活用する	170
4.1.5	A-D変換とファイルフォーマット	171
4.1.6	学習データと評価データ	174
4.2	音響分析と特徴抽出	176
4.2.1	音響分析	176
4.2.2	スペクトル分析	178

4.2.3 MFCC	179
4.2.4 韻律特徴量	181
4.2.5 OpenSMILE	183
4.3 音声認識	184
4.3.1 音声認識システムの構成	184
4.3.2 音声認識のツール	186
4.4 機械学習	186
4.4.1 多変量解析による自動分類	186
4.4.2 多クラス分類問題	188
4.4.3 決定木による分類	189
4.4.4 サポートベクターマシン	191
4.4.5 WEKA	193
4.4.6 ディープラーニング	195
演習問題	197
引用・参考文献	198
演習問題解答	200
索引	207

1 章

音声音響インタフェースの実現のために

◆本章のテーマ

本章では、音を使ったインタフェースの位置づけについて概観し、具体的な例についても学ぶ。そのうえで、音声音響インタフェースの実践のために必要となる基礎知識と、さまざまなツールの活用法について、本書でどのように学んでいくことができるのかについて述べる。

◆本章の構成（キーワード）

- 1.1 身の回りの音声音響インタフェース
スマートフォン、テレビ会議装置
- 1.2 ツールを活用したインタフェース実践
MATLAB, Scilab, openSMILE, Weka

◆本章を学ぶと以下の内容をマスターできます

- ☞ 聴覚の特徴
- ☞ 音声音響インタフェースの実例
- ☞ 音声音響処理に役立つツール

1.1 身の回りの音声音響インタフェース

人間の五感の中でも、とりわけ重要だと思われるのが視覚と聴覚であろう。視覚は、一度に処理できる情報量が多く、空間認知などでは特に重要な役割を担っている。一方、聴覚は言語と深く結びついており、人の精神活動の奥深くにかかわる情報がやりとりされることが多い。

昨今、街には監視カメラが数多く設置されている。以前は嫌がる人も多かったが、犯罪捜査などで有効活用されることも多く、徐々に市民権を得てきているようである。しかし、24時間ずっと録音を続ける装置を、街中に設置したらどうなるだろう。「そんなのは盗聴されているようで嫌だ」という人は多いのではなからうか。言い換えると、音が伝える情報の中には、それだけ機微なものが含まれるということだろう。

産業革命以来の技術革新の歴史の中で、音は重要なテーマの一つであった。音には揮発性と呼ばれる特性があり、音が発せられたその場所・その時間に居合わせない限り、その情報に接することはできなかった。しかし、19世紀後半に電話や蓄音機が発明され、距離や時間に対する制約が取り除かれるようになった。その後、電話の技術は携帯電話に受け継がれ、より効率的に声の情報を伝達するため、さまざまな音声符号化技術が考案されてきた。一方の蓄音機も、コンパクトディスク（CD）からmp3プレーヤーへと変貌していく中で、再生技術や圧縮技術の進化の恩恵を受けている。

このように、「保存」と「伝達」の二つが長年にわたって音声音響処理の中心であったわけだが、近年のIT技術の革新により、「加工」「生成」「理解」といった機能が加わり、総称として「音声インタフェース」「音響インタフェース」という言葉が使われる機会も増えてきた。インタフェースとは、二つのシステムが接する場所で、どのように情報がやりとりされるかを表す言葉であるが、特に人間と機械が接する場所でのやりとりを指して、「ヒューマンマシンインタフェース」という言葉を使うこともある。

実際、われわれの身の回りにはどのような音声音響インタフェースがあるの

か、考えてみよう。

電話は進化してスマートフォンになった。音の強弱を0と1のビット信号で効率的に表すための符号化技術や、帯域を効率的に使うための圧縮技術に加えて、屋外のうるさい環境で使うための雑音抑圧技術や、ハウリングを防ぐためのエコーキャンセラ技術なども、スマートフォンでの通話のためには欠かせない技術である。また、こうした技術は、**テレビ会議装置**などでも役立っている。

音の再生技術は、mp3プレーヤに見られるようなパーソナル化に加えて、映画館やテーマパークで使われる立体化の方向にも進化している。サラウンド再生やバイノーラル再生による臨場感の獲得は、今後はバーチャルリアリティという形でさらに存在感を高めていくと予想される。

音の生成技術は、シンセサイザのような電子楽器に始まり、最近では“**初音ミク**”などの歌声合成システムが人気を集めている。一方で、普通の話し声の合成システムの普及も進み、街中のアナウンスや動画コンテンツ作成などで数多く使われている。

音の理解技術を代表するのは、音声認識システムであろう。長い研究の歴史にもかかわらず、なかなか実用化が進まなかった音声認識だが、スマートフォンの普及により認知度が高まり、少しずつ使われる機会が増えてきている。また、**ビッグデータ**の管理・解析技術の進歩により、コールセンターのログや機器の動作音などを蓄積し、分析していこうという機運も高まっている。

1.2 ツールを活用したインタフェース実践

本書では、こうした音声音響インタフェースを実践的に学習するための材料を提供する。インタフェースを実現するためには、基本原理に加えて、さまざまな周辺機能についての理解が必要である。また、すべての機能を自ら実装するのではなく、ときには既存のツールをうまく活用することにより、作業を効率化することができる。一方、こうしたツールをブラックボックス化すること

4 1. 音声音響インタフェースの実現のために

なく、そこで行われている処理の内容を把握しておくことも、完成したインタフェースを活用する際には重要である。

2章では、音をシステムに取り込み、加工・伝達し、再生するような装置を念頭に置き、そこで用いられるさまざまな処理について解説する。音の信号はさまざまな高さの成分の重ね合わせであり、それを表現するためのフーリエ変換についての理解は不可欠である。ただし、フーリエ変換そのものは、たいいてい場合はツールを使用することにより簡単に実行できるので、変換公式の詳細よりも、むしろその意味についての理解が重要であろう。

エコーキャンセラやマイクロホンアレイなどでは、複数の信号を組み合わせることで所望の信号を得るために、複雑な式を扱うことが求められるが、本書ではそれらの式の詳細を追わなくても、なるべく定性的な理解が可能となるような記述を心がけた。例えば、こうした機能を実行するツールが提供されているときに、どの場面でのどの機能を活用すればよいのか、定性的な検討ができるようになることを、読者には心がけて欲しい。

3章では、信号処理ツールである **MATLAB** あるいは **Scilab** を用いた、具体的な信号処理の実践を学ぶ。Scilab は MATLAB とほぼ同じ機能を持ち、大部分のプログラムは共通である。これらの信号処理ツールはインタープリタ型のツールで、コマンドウィンドウにコマンドを入力することにより、演算や処理が行われる。一連のコマンドをテキストファイルのように記述した、スクリプトと呼ばれるファイルを実行することもできる。また、ツールに用意してある関数以外に、自分で関数ファイルを作成して使用することもできる。

これらのツールの利点は、C 言語や Java のような一般のプログラムに比べ、はるかに短いプログラムで高機能を実現できることである。具体的な演算手順を簡潔明瞭に記述できるため、プログラムのフローチャートとみることもできるし、また、複雑な演算の備忘録にもなる。

MATLAB や Scilab は GUI (graphical user interface) も構築できるほか、マウスクリックによる座標の読み込みなどのインタラクティブな使用が可能であり、本書もその特徴を活かして理解を深められるように工夫した。特に、伝達

関数の極と零点と周波数特性の関係の理解には有効である。

このほか、MATLAB や Scilab には優れた描画機能があり、各種グラフのほか、色表示の機能を活用したサウンドスペクトログラムの表示も含めた。

4章は、特に機械学習という側面に着目した章である。あるテーマのもとにデータを集め、そこから機械学習によりなんらかの知識を獲得し、未知のデータに対する予想を行うという枠組みを自ら実践できるようになることが、この章の目的である。本書では音のデータ処理を扱っているが、ここで述べることの多くは、音以外のメディアデータに対しても役立つはずである。

データ処理や機械学習の分野では、近年さまざまな機能に特化したツールキットが公開されているが、それぞれのツールの使い方を習得すること自体も、決して簡単ではない。そのような中で、特徴抽出ツールの **openSMILE** や、機械学習ツールの **Weka** は汎用性が高く、さまざまな場面で活用することができる。こうしたツールを使いこなして、さまざまな場面で音声音響信号処理の課題にチャレンジできるようになって欲しい。

演 習 問 題

- [1.1] 普段の日常で聴いている音を、「自然のままの音」「自然のものを人間が加工した音」「人間が作った音」に分類してみよう。人間が加工したり作成したりするときには、どういう点に注意しているかを考えてみよう。

【あ】

アクティブノイズコント
 ロール 70
 アップサンプリング
 173, 177
 アナログ-デジタル変換 16
 安定性 107

【い】

位相 9
 一般化カルバック-ライプ
 ラー・ダイバージェンス 59
 インパルス応答
 20, 64, 87, 170
 韻律 181

【う】

ウィナーフィルタ 63

【え】

エイリアシング 17
 エコー 177
 エコーキャンセラ 23
 エコーサプレッサ 23
 枝刈り 191

【お】

オイラーの公式 21, 87
 音のスパース性 49
 オープン評価 174
 音響モデル 174, 185
 音源同定 50
 音源方向推定 48
 音場制御 63
 音声強調 60
 音声スイッチ 24

音素 168

【か】

過学習 191
 可逆圧縮 173
 学習データ 174
 角周波数 9, 86
 カクテルパーティー効果 35
 隠れ層 196
 重ね合わせの原理 8
 可聴域 171
 カーネルトリック 192
 カラーマップ 135
 カルバック-ライブラー・
 ダイバージェンス 56

【き】

帰還 101
 揮発性 166
 基本周波数 150, 182
 逆位相 70
 極 96

【く】

空間的エイリアシング 36
 クローズド評価 174
 クロスバリデーション 175

【け】

決定木 190
 ケプストラム 149, 150
 言語モデル 185

【こ】

公開データ 170
 効果音 123
 高速フーリエ変換 18, 139, 179

誤差関数 26
 5.1 チャンネルサラウンド 66
 混合ガウス分布 195
 コンボリューション 89

【さ】

最急降下法 28
 最小二乗誤差アルゴリズム 28
 最小分散ビームフォーマ 48
 Scilab 4
 サイン波 9
 サウンドスペクトログラム
 131
 雑音抑圧 36, 60
 サポートベクター 191
 サポートベクターマシン 191
 サラウンド 66
 三角窓 133
 残響 128
 残差相関法 151
 サンプリング 16
 サンプリング周波数
 16, 78, 171
 サンプル 16

【し】

死角形成型ビームフォーマ 44
 時間領域 86
 指向性マイクロホン 42
 自己相関行列 27
 自己相関係数 141
 事前 SNR 62
 遮断周波数 85
 収束する 103
 周波数 9
 周波数変化音 123
 周波数変調 74, 124

- | | | | | | |
|-------------------|----------|---------------------|------------|--------------------|----------|
| 周波数領域 | 86 | 畳み込み定理 | 20 | トランスオーラル再生 | 68 |
| 出力層 | 196 | タップ数 | 20 | | |
| 瞬時角周波数 | 123 | ダブルトーク検出 | 33 | 【な】 | |
| 瞬時周波数 | 123 | 多変量解析 | 187 | ナイキスト周波数 | 17 |
| シンセサイザ | 155 | ダミーヘッド | 67 | | |
| 深層学習 | 195 | 多様性 | 168 | 【に・の】 | |
| 振幅 | 9 | 短時間スペクトル | 14 | 入力層 | 196 |
| 振幅変調 | 74 | 単振動 | 9 | ニューラルネットワーク | 196 |
| | | | | ノード | 190 |
| 【す】 | | 【ち】 | | 【は】 | |
| スケーリングの不定性 | 57 | チェビシェフ | 118 | 葉 | 190 |
| ステレオ再生 | 65 | 遅延和ビームフォーマ | 39 | バイナリマスキング | 54 |
| スピーカアレイ | 72 | 調音結合 | 168 | バイノーラル録音 | 67 |
| スプライン関数 | 125 | 重畳 | 169 | ハイレゾ音源 | 172 |
| スペクトル | 13 | 重畳積分 | 89 | ハウリング | 23 |
| スペクトルサブトラク
ション | 60 | 調波構造 | 182 | 白色雑音 | 119 |
| スペクトル包絡 | 145 | 直流成分 | 12 | 波形編集ソフト | 177 |
| スペクトログラム | 15 | | | バターワース | 118 |
| スマートフォン | 3 | 【つ】 | | バックプロパゲーション | 196 |
| | | 通過帯域幅 | 85 | 発散する | 107 |
| 【せ】 | | 【て】 | | 話し言葉 | 169 |
| 正帰還 | 101 | 低周波発振器 | 159 | ハニング窓 | 133 |
| 声調言語 | 181 | ディープニューラル
ネットワーク | 196 | パーミュテーションの
不定性 | 57 |
| 零点 | 96 | ディープラーニング | 186, 195 | ハミング窓 | 133 |
| 線形エコーキャンセラ | 31 | 適応ノイズキャンセラ | 38 | パラメトリックスピーカ | 73 |
| 線形判別分析 | 188 | 適応ビームフォーマ | 46 | 半二重 | 23 |
| 線形予測係数 | 139 | デコーダ | 185 | | |
| 線形予測分析 | 139 | デシベル (dB) | 82 | 【ひ】 | |
| 全二重 | 23 | Toeplitz 型行列 | 142 | 非可逆圧縮 | 173 |
| | | デルタ関数 | 87, 89 | 非線形エコーキャンセラ | 31 |
| 【そ】 | | テレビ会議装置 | 3 | ビッグデータ | 3, 186 |
| 相関係数 | 55 | 伝達関数 | 20, 85, 87 | ピッチ抽出 | 150 |
| 相互相関ベクトル | 27 | | | 非同期分散マイクロホン
アレイ | 52 |
| 増幅 | 177 | 【と】 | | 非負値行列因子分解 | 59 |
| ソフトマージン SVM | 192 | 統計的雑音抑圧 | 62 | ビブラート | 123, 124 |
| 疎密波 | 7 | 頭部伝達関数 | 68 | ビームフォーマ | 39 |
| | | 特定話者システム | 174 | ビームフォーミング | 39 |
| 【た】 | | 独立性 | 55 | 評価データ | 174 |
| ダイナミックレンジ | 172 | 独立成分分析 | 57 | 標本化定理 | 17 |
| ダウンサンプリング | 172, 177 | トライフォン | 168 | | |
| 畳み込み演算 | 20 | | | | |

【ふ】

フィルタ 20
 フィルタアンドサム
 ビームフォーマ 44
 フェードアウト 177
 フェードイン 177
 フォルマント 182
 負帰還 101
 節 190
 不特定話者システム 174
 ブラインド信号分離 53
 ブラックマン窓 133
 フーリエ級数 11, 12, 87
 フレーム 15, 179
 フレーム処理 15
 ブロッキング行列 47

【へ】

平面波近似 39
 変形相関法 151
 変調周波数 124

【ほ】

ボコーダ 139

【ま】

マイクロホンアレイ 35
 マージン最大化 191
 MATLAB 4
 窓掛け 14
 窓関数 132, 133

【み】

ミキサー 65
 ミキシング 65

【む】

無響室 167

【め】

メルケプストラム係数 179

【ゆ】

ユークリッド距離 154
 Yule-Walker 方程式 142

【よ】

予測誤差 139
 予測残差 151

読み上げ音声 169

【ら】

ラプラス変換 87

【り】

リバーブ 123, 177
 リーフ 190
 量子化ビット数 172
 両耳間強度差 36
 両耳間時間差 36

【る】

類似度 153

【ろ】

録音環境 167
 録音装置 166
 ロンバード効果 170

【わ】

WAV 形式 173

【A ~ G】

A-D 変換 16
 ARMA フィルタ 107
 A 特性フィルタ 181
 CPU 34
 DTMF 178
 DTW 154
 ERLE 28
 FFT 18, 132
 FIR フィルタ 92
 GCC-PHAT 49
 Griffith-Jim 型適応ビーム
 フォーマ 46

【H ~ N】

HRTF 68
 ICA 57
 IID 36
 IIR フィルタ 92
 ITD 36
 LMS アルゴリズム 28
 LPC 139
 MFCC 154, 179
 NLMS 30
 NMF 59

【O ~ Z】

openSMILE 5
 PCM 173
 Q 値 85
 SN 比 170
 SPL 83
 SVM 191
 TSP 信号 64
 Weka 5
 z 変換 88

— 著者略歴 —

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

大淵 康成 (おおぶち やすなり)

1988年 東京大学理学部物理学科卒業
1990年 東京大学大学院理学系研究科修士課程修了(物理学専攻)
1992年 株式会社日立製作所中央研究所勤務
2002年 カーネギーメロン大学客員研究員
~03年 兼務
2005年 早稲田大学客員研究員兼務
~10年
2006年 博士(情報理工学)(東京大学)
2013年 クラリオン株式会社勤務(兼務)
~15年
2015年 東京工科大学教授
現在に至る

音声音響インタフェース実践

Practical Speech and Sound Interfaces

© Kiyooki Aikawa, Yasunari Obuchi 2017

2017年3月13日 初版第1刷発行



検印省略

著者 相川 清明
大淵 康成
発行者 株式会社 コロナ社
代表者 牛来真也
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 978-4-339-02793-8

(高橋)

(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