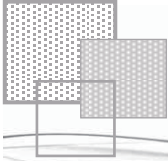


# MATLABで学ぶ 実践画像・音声処理入門

博士(工学) 伊藤 克亘  
工学博士 花泉 弘 共著  
博士(工学) 小泉 悠馬

コロナ社



# まえがき

音声や画像は、コンピュータ処理の対象となって、メディアとしての可能性が飛躍的に広がりました。そのようなデジタル処理技術の多くは、数学に基づいています。また、音声や画像は物理的な現象なので、物理モデルも多用されます。物理モデルの多くも数式に基づいています。

MATLAB は、それらの処理を、ほぼ数式をそのまま記述するのと同様の形式でプログラミングできます。そのため、論文など専門文書で数式に基づいて紹介される手法を直接試すのに便利です。

本書では、そういった手法を身につけるための基本的な手法を簡単な数式に基づく処理のプログラムとして取り上げています。これらのプログラムで、実際の音声データや画像データを処理しながら学ぶことを目的としています。コンピュータによる音声や画像の簡単な加工、分析、生成方法を、音声処理と画像処理の共通性を意識しながら学べるような内容にしました。

## 対象読者

本書では、数式に基づく手法の説明に集中するため、MATLAB のプログラミング言語としての文法や使い方などは、必要最低限しか取り上げません。幸い、プログラミング言語としての MATLAB の使い方については、MathWorks 社の MATLAB トレーニングのページ (<https://matlabacademy.mathworks.com/jp>) にある無料コースの「MATLAB 入門」が優れています。本書では、このコースを修了していることを想定しています。

数学に関しては、高校数学および大学レベルの微積分、線形代数、統計学の

基礎知識があることを想定しています。またフーリエ変換については、どのようなものか知っている方がサンプルプログラムを理解しやすいでしょう。数学に関しては、必要に応じて、これらのレベルの基礎的な教科書を参照すれば十分でしょう。

## 本書の構成

本書では、説明内容に合わせて、数行のプログラムが示されます。これらのプログラムは、章内では、全部続けて実行することを想定しています。つまり、章の最初の方で値を設定された変数が、後のプログラムで断りなしに使われることもあります。

また、章の最初にキーワードを示します。それに関連する数学的な事項を思い出すとよいでしょう。

その章で学んだことを定着させるための章末問題も用意しています。これらの問題は、学んだ知識を自分なりに応用するためのヒントになっています。定着させるためには、最後まで自力で考えた方がよいのですが、巻末にはヒントを掲載しました。

## 本書で想定する MATLAB 環境

本書のサンプルプログラムは、Windows 10 上の MATLAB 2019a の環境で動作確認しています。Mac や Linux など異なる環境では、出力などが多少異なることもありうるのでご注意ください。

本書で用いた音声や画像データ、本書のために開発したパッケージもサポートサイト ([www.coronasha.co.jp/np/isbn/9784339009255/](http://www.coronasha.co.jp/np/isbn/9784339009255/)) からダウンロードできます。

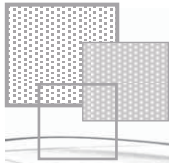
本書では、MATLAB 本体以外につぎの Toolbox を用います。Image Pro-

cessing Toolbox (2, 5, 6, 10~12 章), Signal Processing Toolbox (3~5, 7~9, 12 章), Statistics and Machine Learning Toolbox (6, 7, 11, 12 章)。

科学技術プログラミングの分野では、MATLAB と同様な用途で、Python の NumPy などのモジュールが使われます。拙著『Python で学ぶ実践画像・音声処理入門』（2018 年，コロナ社刊）は本書とほぼ同じ内容となっているので、Python のプログラムを MATLAB に移植したり，その逆の用途にも役立つと思います。これらの書籍が，音声・画像処理の実践的なプログラミングを習得する一歩になれば幸いです。

2019 年 7 月

著 者



# 目 次

## 1

### 簡単な音声処理

1.1	波形データの生成	1
1.2	1次元データの可視化	5
1.3	時間波形の重ね合わせ	9
1.4	時間波形の連結	12
1.5	読み込んだ音声データの加工	13
	章末問題	17

## 2

### 簡単な画像処理

2.1	画像の構造	19
2.2	画像・ビデオの読み込み	23
2.3	領域の抽出	28
	章末問題	34

## 3

### 音声のフーリエ変換

3.1	周期現象	36
3.2	フーリエ変換	37
3.3	窓関数	40

3.4 実データのスペクトル解析 .....	44
3.5 スペクトログラム .....	47
3.6 逆フーリエ変換 .....	50
章 末 問 題 .....	52

## 4 フィルタ (音声)

4.1 線形フィルタ .....	55
4.1.1 線形システム .....	55
4.1.2 遅延演算 .....	56
4.1.3 移動平均フィルタ .....	58
4.2 インパルス応答 .....	61
4.3 IIR フィルタ .....	63
4.4 フィルタ設計のツール .....	65
章 末 問 題 .....	66

## 5 画像の周波数領域処理

5.1 空間周波数 .....	69
5.2 2次元フーリエ変換 .....	71
5.3 周波数領域でのフィルタ処理 .....	74
5.4 周波数領域での画像の拡大 .....	77
章 末 問 題 .....	78

## 6 画像の空間領域処理

6.1 2次元畳み込み .....	80
-------------------	----

6.2 微分演算	84
6.3 エッジの検出	85
6.4 非線形フィルタ	89
章末問題	89

## 7 音声データの相関

7.1 相互相関	91
7.1.1 ベクトルの類似度	91
7.1.2 相互相関関数	95
7.2 自己相関	96
7.3 時間波形のフレーム処理	99
章末問題	104

## 8 画像データの類似度

8.1 画素のユークリッド距離	106
8.2 画素の相関の応用	107
8.3 領域の相関	109
章末問題	113

## 9 複素信号

9.1 信号の複素指数関数表現	115
9.2 周波数変調	117
9.2.1 瞬時周波数	117
9.2.2 周波数変調	118

9.2.3 任意の音の周波数変調 .....	119
章 末 問 題 .....	123

**10****画像の幾何学的処理**

10.1 2次元平面上の回転 .....	126
10.2 2次元平面上の平行移動 .....	128
10.3 同次座標表現を用いた変換 .....	129
10.4 アフィン変換 .....	130
10.5 射影変換 .....	133
10.6 複雑な形状の変換 .....	133
章 末 問 題 .....	136

**11****分 類**

11.1 特 徴 量 .....	138
11.1.1 短時間エネルギー .....	139
11.1.2 零 交 差 .....	140
11.2 k最近傍分類 .....	142
11.3 最 尤 法 .....	146
章 末 問 題 .....	148

**12****音声・画像処理の応用**

12.1 Wavetable 合成 .....	150
12.1.1 ADSR エンベロープ .....	150
12.1.2 楽器音からの波形データの抽出 .....	153

12.1.3	複数のテンプレートを用いた合成	154
12.1.4	長さの変更	157
12.1.5	リサンプルによるピッチの変更	157
12.2	衛星画像の時間変化領域の解析	158
章末問題		163
<b>章末問題ヒント</b>		166
<b>索引</b>		185

# 1

# 簡単な音声処理

まず、1次元のデジタルデータの代表として、音データを取り上げる。この章では、コンピュータに読み込んだ音声データを用いて、MATLABによる1次元データの処理方法を学ぶ。

**キーワード** 振幅, 周波数, サンプリング周期, サンプリング周波数, 離散的, スカラ, ベクトル, 可視化

## 1.1 波形データの生成

最も単純な音の一つに**純音**がある。純音は**正弦波**で表される。物理の教科書を見ると、純音は式 (1.1) のように表される。

$$y = A \sin(2\pi ft) \tag{1.1}$$

ここで、 $y$  は音圧、 $A$  は**振幅**、 $f$  は**周波数**、 $t$  は時間である。 $y$  の変化のグラフが図 1.1 である。

グラフは2次元であるが、 $y$  という一つの変数が各時刻で一つの値をとり、その値が時間で変化するので1次元データと呼ぶ。

式 (1.1) に基づいて、MATLAB で音のデジタルデータを生成するプログラムを作成する。データを生成するためには、まず、式 (1.1) の変数の値を決めなければならない。 $A$ 、 $f$  は、それぞれ一つの値を定めればよい。例えば、 $A = 1$ 、 $f = 50$  などである。このように大きさのみを持つ量のことを**スカラ**と呼ぶ。

一方、 $t$  については、例えば、1秒間のデータを作成するときには、開始時刻を

## 2 1. 簡単な音声処理

0秒とすると、時刻0秒から1秒まで変化する。式で書くと  $0 \leq t \leq 1$  となる。

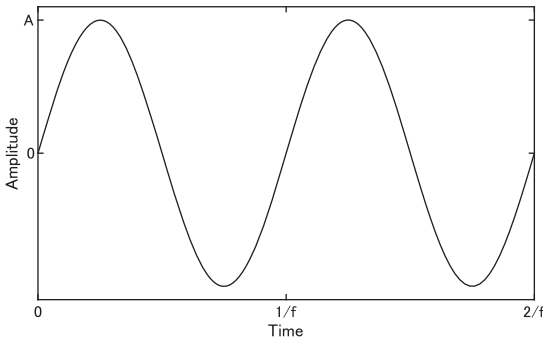


図 1.1 正弦波のグラフ

つまり一つの値ではない。

物理的な世界では、 $t$ は連続的に変化する。このような連続的な量をコンピュータで扱う最も一般的な方法は、均等で微小な間隔の値の列として表現することである。例えば、 $1/8000$ 秒の間隔とす

ると、 $0 \leq t \leq 1$  という範囲は、 $0, 1/8000, 2/8000, 3/8000, \dots, 7999/8000, 1$  という数の列で表される。この間隔  $1/8000$  秒を**サンプリング周期**と呼ぶ。また、この周期に対応する周波数を**サンプリング周波数**と呼ぶ。周波数は周期の逆数なので、この場合は  $8000 \text{ Hz}$  となる。

このようにとびとびの値で表すことを**離散的**であるという。つまり、コンピュータの中では正弦波は、**図 1.2** の丸印の値だけで表される。

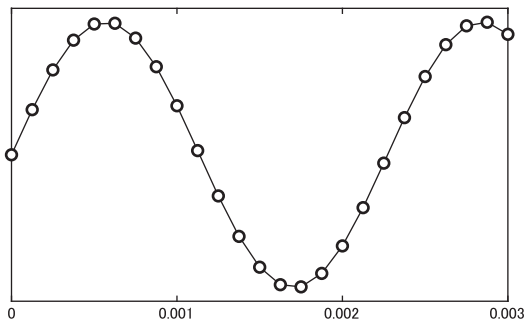


図 1.2 離散的な正弦波データ

このようなデータをプログラミングするときには、 $t$ をひとまとまりで扱うと便利である。そのような場合、数学的には**ベクトル**として扱うことがある。プログラムでは、ベクトルをひとまとまりで扱えると便利である。

● 正弦波の生成 MATLAB で 440 Hz の音を 1 秒間生成してみる。コマンドウィンドウにプログラム 1-1 のように入力する。「>>」は、MATLAB が表示するプロンプトである。したがって試すときには、入力しなくてよい。

プログラム 1-1 (正弦波の生成)

```
>> fs=8000;
>> t=0:1/fs:1; ❶
>> f=440;
>> a=0.8;
>> y=a*sin(2*pi*f*t); ❷
```

このように、MATLAB はベクトルを用いた計算を非常に簡単に、ほとんど数式と同じ形でプログラミングできる。❶では、時間を表す数列を作成している。「:」は、MATLAB ではいろいろな意味を持つ記号であるが、ここでは、数列を生成するために利用されている。

数列を生成するときに使う「:」の使い方を示す。

```
>> n=0:10
n =
    0    1    2    3    4    5    6    7    8    9   10
```

このように、「開始値：終了値」という形で指定する（この例の場合、開始値は 0、終了値は 10）と、0 から 10 までの整数列が生成され、変数  $n$  に代入される（MATLAB では、変数自体には型がないことに注意）。

プログラム 1-1 の  $t$  の例は「:」が二つ使われている。この場合は、「開始値：間隔：終了値」となる。間隔を指定した「:」の例を示す。

```
>> m=-1:0.5:2
m =
   -1.0000  -0.5000   0   0.5000   1.0000   1.5000   2.0000
```

指定された通りに 0.5 刻み間隔で -1 から 2 までの値からなる数列が生成される。行の最後の「;」は、値の表示を抑制する。つぎに例を示す。

```
>> a=1 ❶
a =
    1
>> b=2; ❷
>> c=a+b; ❸
```

❶のように、代入など計算を行うと値が表示される。しかし、❷のように「;」を付けるとなにも表示されない。❸のように計算したときも、その式に「;」を付けると結果は表示されない。

プログラム 1-1 の❷は、式 (1.1) を MATLAB でプログラミングしている。数式とほとんど同じ形で表現できることに注目してほしい。この `pi` は、MATLAB であらかじめ用意されている変数で  $\pi$  の値を持つ。

`sin` は、三角関数の `sin` である。MATLAB で用意されている関数については、`help` コマンドで説明を見られる。

```
>> help sin
sin - ラジアン単位の引数の正弦

この MATLAB 関数 は、X 要素の正弦を返します。

Y = sin(X)

参考 asin, asind, sind, sinh

sin のリファレンスページ
sin という名前のその他の関数
```

`sin` のリファレンスページの部分をクリックするとさらに詳しい説明が読める。`help` の説明では関数名が大文字になっているが、実際に使うときには小文字でないといけないことに注意せよ。この説明では、引数に関して「X 要素の」と書かれている。プログラム 1-1 の 4 行目を見ればわかるように、この `sin` は引数に数列（配列）をとれる。引数に数列をとった例を示す。

```
>> sin(0:0.1:1)
ans = ①
Columns 1 through 7
    0    0.0998    0.1987    0.2955    0.3894    0.4794    0.5646
Columns 8 through 11
    0.6442    0.7174    0.7833    0.8415
```

① の `ans` は、直前の計算結果が自動的に代入される変数である。この例のように、`sin` は、数列を引数に与えると、計算結果も数列となり、その値は引数の `sin` の値である。このように、MATLAB の多くの関数は数列や行列を引数にとることができ、数式とほぼ同じ形でプログラミングできる。

プログラム 1-1 の ② では、`t` は数列である。`t` に掛けられている  $2\pi f$  は  $\pi$ 、 $f$  がスカラなので、スカラである。MATLAB では数列をベクトルとして扱う。また、MATLAB では数式同様ベクトルにスカラが掛けられる場合は、ベクトルのそれぞれの要素をスカラ倍する。したがって、 $2\pi f t$  は、数列  $t$  のおのおのの要素をそれぞれ  $2\pi f$  倍する。すべての要素が  $2\pi f$  倍された数列に対して `sin` を計算し、その結果の数列を  $a$  倍する。

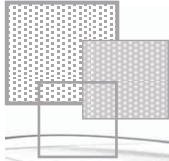
作成した音データを出力するための関数が `sound` である。プログラム 1-1 で作成した `y` は、つぎのようにして出力する。

```
>> sound(y, 8000)
```

第1引数が出力したい数列、第2引数はサンプリング周波数である。周波数が440 Hz の「ラ」の音が1秒間聞こえるはずである。

## 1.2 1次元データの可視化

音を生成したり、加工したり、録音した場合には、もちろん、音を再生、出力して確認すべきである。しかし、音は聞こえ方が人によってかなり異なるし、プログラムに失敗していたら、聞こえる音にならなかったり、デバイスに悪影



# 索引

<b>【あ】</b>		カテゴリ	142		
アフィン変換	131	カラーエッジ	107	<b>【し】</b>	
<b>【い】</b>		<b>【き】</b>		時間解像度	50
位相	40	幾何学変換オブジェクト	130	時間波形	6
一次式	56	基本周波数	98	時間領域	50
移動平均フィルタ	58	逆フーリエ変換	51	しきい値	30
イメージシーケンス	27	教師付き分類	142	システム	55
インデクス	6, 8	教師なし分類	159	射影変換	133
インパルス	61	行ベクトル	11	周期関数	96
インパルス応答	61	距離	91	周期信号	36
<b>【う】</b>		<b>【く】</b>		周波数	1
うなり	10	空間周波数	70	周波数応答	61
<b>【え】</b>		空間周波数スペクトル	70	周波数特性	61
衛星画像	144	矩形波	52	周波数分解能	39
エッジ	70	クラスタリング	159	周波数変調	118
エッジ検出	87	クリップ	12	周波数領域	50
エネルギー	139	グレイスケール画像	19	縮小	129
<b>【お】</b>		<b>【こ】</b>		出力	55
オイラーの公式	115	コサイン類似度	93	純音	1
オブジェクト	27	<b>【さ】</b>		瞬時周波数	118
音声波形	6	斉次座標表現	129	シンセサイザー	150
<b>【か】</b>		最小値	32	振幅	1
回転	126	最大値	32	振幅スペクトル	38
角周波数	117	最短距離法	149	振幅変調	12
拡大	129	最尤法(分類)	147	<b>【す】</b>	
角度	92	座標	20	スカラ	1
加算	55	サンプリング周期	2, 56	スペクトル	36
可視化	6	サンプリング周波数	2	スペクトログラム	50
				<b>【せ】</b>	
				正規化	12
				正規化相関法	149
				正弦波	1

零交差	140			ベクトル	2
鮮鋭化	87			変換行列	127
漸化式	64				
線形(線型)システム	56			<b>【ほ】</b>	
線形チャープ	117			補間	121
線形補間	121			ホワイトノイズ	58
せん断	136			<b>【ま】</b>	
		<b>【の】</b>		マスク	29
		ノコギリ波	53	窓関数	42
		ノルム	92		
		<b>【は】</b>		<b>【め】</b>	
		倍音構造	37	メディアン	89
		バイナリマスク	29		
		白色雑音	58	<b>【ゆ】</b>	
		外れ値	110	床関数	53
		ハニング窓	42	ユークリッド距離	91
		反転	136		
		バンド	21	<b>【ら】</b>	
		ハン窓	42	ラプラシアンオペレータ	87
				乱数	58
		<b>【ひ】</b>		<b>【り】</b>	
		非線形フィルタ	89	リーク	42
		ビブラート	118	離散的	2
		微分係数	84	離散フーリエ変換	37
		標準正規分布	58		
		<b>【ふ】</b>		<b>【る】</b>	
		フィードバック	64	類似度	91
		フィルタ	58	<b>【れ】</b>	
		複素共役	51	列ベクトル	14
		複素指数関数	115	<b>【ろ】</b>	
		複素数平面	40	ローパスフィルタ	65
		符号関数	141	論理インデクス	29
		フーリエ変換	36	論理積	81
		フレーム	49		
		フレーム(ビデオ)	26		
		フレームシフト	51		
		フレームレート	27		
		プロット	6		
		プロパティ	26		
		分散	147		
		分類	138		
		分類器	144		
		<b>【へ】</b>			
		平滑化	67		
		平均	53, 147		
		平行移動	128		
<b>【そ】</b>					
相関係数	94				
相関係数行列	108				
相互相関関数	95				
		<b>【た】</b>			
帯域	65				
ダイナミックレンジ	110				
多次元正規分布	147				
畳み込み	60				
多バンド画像	159				
単位インパルス	61				
短時間エネルギー	139				
		<b>【ち】</b>			
遅延	56				
チャープ信号	117				
中央値	89				
直流成分	71				
		<b>【て】</b>			
定数倍	56				
デローネイ三角分割	133				
		<b>【と】</b>			
同次座標表現	129				
特徴空間	106, 142, 147				
特徴量	138				
		<b>【な】</b>			
ナイキスト周波数	38				
内積	92				
		<b>【に】</b>			
入力	55				

<b>[A]</b>	
abs	37
ADSR	151
affine2d	130
angle	40
audioread	13
audiowrite	16
axis	76, 85, 127
<b>[B]</b>	
bitand	81
butter	66
<b>[C]</b>	
caldays	67
ceil	79
chirp	118
circshift	57
ClassificationKNN	144
colormap	85
conj	51
conv	60
conv2	80
corrcoef	108
cpcorr	137
cpselect	137
csvread	67 135
<b>[D]</b>	
datetime	67
deLaunay	133
diff	53
dot	92
<b>[E]</b>	
end	8
exp	116
<b>[F]</b>	
$f_0$	98
fft	37
fftshift	72

figure	50
filter	64
filter2	81
find	39
FIR	64
fir1	65
fitcknn	143, 144
fitgeotrans	132
flipud	51
floor	53
FrameRate	26
fspecial	83
<b>[H]</b>	
hann	42
hasFrame	27
help	4
histeq	135
hold	43
HPF	68
hypot	86
<b>[I]</b>	
ifft	51
ifft2	75
IIR	64
iirnotch	68
imcrop	24
imnoise	90
implay	27
imread	23
imrotate	128
imshowpair	135
imtool	20
imwarp	130
imwrite	25
interp1	121
intmax	22
inv	132
isempty	103
islocalmax	46

<b>[K]</b>	
k 最近傍分類	143
k 平均法	159
kmeans	149, 159
<b>[L]</b>	
length	8
linspace	20
load	157
lookfor	15
LPF	65
<b>[M]</b>	
max	32
mean	53
medfilt2	89
mesh	73
meshgrid	75
min	32
mvnpdf	147
<b>[N]</b>	
nan	103
nanmean	103
nargin	111
nextpow2	53
norm	92
normxcorr2	113
<b>[O]</b>	
ones	21
<b>[P]</b>	
pi	4
plot	6
poissrnd	84
predict	144
Prewitt オペレータ	89
<b>[R]</b>	
randi	33
randn	58

rats	40	square	153		
readFrame	27	std	109	<b>[X]</b>	
repmat	22	stem	37	xcorr	95
resample	105	strcmp	144	xcorr2	112
reshape	20, 33	subplot	57	xlabel	7
RGB	20	sum	92	<b>[Y]</b>	
RGB 色空間	106			ylabel	7
rgb2gray	70	<b>[T]</b>		<b>[Z]</b>	
<b>[S]</b>		tic	108	zeros	21
save	156	toc	108	<b>[記号・数字]</b>	
sawtooth	104	tripplot	128	\	132
scatter	110	<b>[U]</b>		'	15, 71
setdiff	143	uint8	20	...	31
sign	52, 141	<b>[V]</b>		:	3
sign	52	vecnorm	108	;	3, 23
sin	4	VideoReader	26	2 階微分	85
size	14, 24	view	73	2 次元逆フーリエ変換	75
Sobel オペレータ	86	<b>[W]</b>		2 次元畳み込み	80
sort	161	Wavetable 合成方式	150	2 次元フーリエ変換	72
sound	5			8 ビット符号なし整数	20
soundsc	12				
spectrogram	49				

— 著 者 略 歴 —

伊藤 克亘 (いとう かつのぶ)

1993年 東京工業大学大学院理工学研究科博士  
課程修了 (情報工学専攻)  
博士 (工学)  
1993年 電子技術総合研究所研究員  
2003年 名古屋大学大学院助教授  
2006年 法政大学教授  
現在に至る

花泉 弘 (はないずみ ひろし)

1981年 東京大学大学院工学系研究科博士課程  
中退 (計数工学専攻)  
1981年 東京大学助手  
1987年 工学博士 (東京大学)  
1987年 法政大学専任講師  
1989年 法政大学助教授  
1996年 法政大学教授  
現在に至る

小泉 悠馬 (こいずみ ゆうま)

2014年 法政大学大学院情報科学研究科博士前  
期課程修了 (情報工学専攻)  
2014年 日本電信電話株式会社 NTT メディア  
インテリジェンス研究所研究員  
現在に至る  
2017年 電気通信大学大学院情報理工学研究科  
博士後期課程修了 (情報学専攻)  
博士 (工学)

## MATLAB で学ぶ実践画像・音声処理入門

Introduction to Media Computing in MATLAB : A Practical Approach

© Katsunobu Ito, Hiroshi Hanaizumi, NTT 2019

2019年9月30日 初版第1刷発行



検印省略

著 者 伊 藤 克 亘  
花 泉 弘  
小 泉 悠 馬  
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 <http://www.coronasha.co.jp>

ISBN 978-4-339-00925-5 C3055 Printed in Japan

(三上)



 <出版者著作権管理機構 委託出版物>

本書の無断複製は著作権法上での例外を除き禁じられています。複製される場合は、そのつど事前に、出版者著作権管理機構 (電話 03-5244-5088, FAX 03-5244-5089, e-mail: info@jcopy.or.jp) の許諾を得てください。

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。