

日本音響学会 編
音響テクノロジーシリーズ 22

音声分析合成

博士（工学） 森勢 将雅 著

コロナ社

音響テクノロジーシリーズ編集委員会

編集委員長

千葉工業大学
博士（工学） 飯田 一博

編集委員

東北学院大学
博士（情報科学） 岩谷 幸雄

甲南大学
博士（情報科学） 北村 達也

滋賀県立大学
博士（工学） 坂本 眞一

八戸工業大学
博士（工学） 三浦 雅展

千葉工業大学
博士（工学） 大川 茂樹

東京大学
博士（工学） 坂本 慎一

神戸大学
博士（工学） 佐藤 逸人

（五十音順）

（2017年11月現在）

発刊にあたって

音響テクノロジーシリーズは1996年に発刊され、以来20年余りの期間に19巻が上梓された。このような長期にわたる刊行実績は、本シリーズが音響学の普及に一定の貢献をし、また読者から評価されてきたことを物語っているといえよう。

この度、第5期の編集委員会が立ち上がった。7名の委員とともに、読者に有益な書籍を刊行し続けていく所存である。ここで、本シリーズの特徴、果たすべき役割、そして将来像について改めて考えてみたい。

音響テクノロジーシリーズの特徴は、なんといってもテーマ設定が問題解決型であることであろう。東倉洋一初代編集委員長は本シリーズを「複数の分野に横断的に関わるメソッド的なシリーズ」と位置付けた。従来の書籍は学問分野や領域そのものをテーマとすることが多かったが、本シリーズでは問題を解決するために必要な知見が音響学の分野、領域をまたいで記述され、さらに多面的な考察が加えられている。これはほかの書籍とは一線を画するところであり、歴代の著者、編集委員長および編集委員の慧眼の賜物である。

本シリーズで取り上げられてきたテーマは時代の最先端技術が多いが、第4巻「音の評価のための心理学的測定法」のように汎用性の広い基盤技術に焦点を当てたものもある。本シリーズの役割を鑑みると、最先端技術の体系的な知見が得られるテーマとともに、音の研究や技術開発の基盤となる実験手法、測定手法、シミュレーション手法、評価手法などに関する実践的な技術が修得できるテーマも重要である。

加えて、古典的技術の伝承やアーカイブ化も本シリーズの役割の一つとなる。例えば、アナログ信号を取り扱う技術は、技術者の高齢化により途絶の危

機にある。デジタル信号処理技術がいかに進んでも、ヒトが知覚したり発したりする音波はアナログ信号であり、アナログ技術なくして音響システムは成り立たない。原理はもちろんのこと、ノウハウも含めて、広い意味での技術を体系的にまとめて次代へ継承する必要があるだろう。

コンピュータやネットワークの急速な発展により、研究開発のスピードが上がり、最新技術情報のサーキュレーションも格段に速くなった。このような状況において、スピードに劣る書籍に求められる役割はなんだろうか。それは上質な体系化だと考える。論文などで発表された知見を時間と分野を超えて体系化し、問題解決に繋がる「メソッド」として読者に届けることが本シリーズの存在意義であるということ再認識して編集に取り組みたい。

最後に本シリーズの将来像について少し触れたい。そもそも目に見えない音について書籍で伝えることには多大な困難が伴う。歴代の著者と編集委員会の苦労は計り知れない。昨今、書籍の電子化についての話題は尽きないが、本文の電子化はさておき、サンプル音、説明用動画、プログラム、あるいはデータベースなどに書籍の購入者がネット経由でアクセスできるような仕組みがあれば、読者の理解は飛躍的に向上するのではないだろうか。今後、検討すべき課題の一つである。

本シリーズが、音響学を志す学生、音響の実務についている技術者、研究者、さらには音響の教育に携わっている教員など、関連の方々にとって有益なものとなれば幸いである。本シリーズの発刊にあたり、企画と執筆に多大なご努力をいただいた編集委員、著者の方々、ならびに出版に際して種々のご尽力をいただいたコロナ社の諸氏に厚く感謝する。

2018年1月

音響テクノロジーシリーズ編集委員会
編集委員長 飯田 一博

ま え が き

音声合成の研究というと、多くの人はテキストを読み上げる text-to-speech (TTS) の研究を想像するかもしれない。本書のタイトルを音声分析合成としたのは、音声合成というタイトルによって、TTS の技術解説を期待する読者が肩透かしを食うことへの懸念からである。音声分析合成とは、音声をなんらかのパラメータとして表現し、表現されたパラメータから音声波形を生成する信号処理技術の総称である。また、一連の技術を束ねたシステムを音声分析合成システムと呼称する。音声の読み上げや加工においては、音声パラメータの生成や加工により実現するアプローチが幅広く利用されている。一方、近年では、WaveNet のように音声波形そのものを出力対象とする技術が実用化され、音声をパラメータで表現する音声分析合成の重要度は相対的に下がったといえる。本書を執筆しているいまも、TTS に関する研究は日進月歩の進歩によるパラダイムシフトを迎えつつあり、10 年後には本書の内容も古典的なものとして扱われるかもしれない。それは音声処理に関する技術のブラックボックス化に繋がりますが、音声を扱うための信号処理技術を習得することは、今後も重要な価値があると筆者は信じている。

本書の目標は、これから音声分析合成システムの研究、あるいはシステムを活用した研究をしたい読者に必要となる知識が、この 1 冊を読むことで一通り習得できることである。音声分析合成システムを構成する信号処理理論については、「高品質音声分析合成に関する信号処理理論の理解」をゴールに定め、そのゴールに必要な数学的知識に限定して紹介する。歴史的に議論を避けることができない伝統的なアルゴリズムは解説するが、さまざまな内容を網羅的に紹介する辞書的な使い方ができる教科書とは位置付けが異なる。本書の想定する具

体的な読者層は、大学で学ぶレベルの微分積分や線形代数の知識に加え、デジタル信号処理を活用した音の信号処理について最低限の知識を習得している大学院生や若手研究者である。

音声分析合成の研究には、信号処理に関する数学的な知識を習得するだけではなく、ときには自分自身で音声を収録することもあり、音声の品質評価法を習得することも不可欠であろう。本書には、高品質な音声を入手する必要に迫られた読者が、最低限必要となる品質で音声を収録する際に助けとなる情報も含むこととした。収録音声に関する条件や収録環境に関する知識は、音声分析や合成時において予期せぬエラーが生じた際に役立つこともある。

本書を読み解くのに必要な数学や、音声収録に関する知識は1章にまとめた。2章では、音声分析合成に特化した音声信号のモデル化と音声分析合成の歴史的な技術を概説する。これらは雑多な内容を広く浅く扱った章であるため、すでに関連知識を有する読者はスキップしてもさしつかえない。3~5章は、2章で説明するボコーダに関連した音声パラメータ群の推定法について、基盤から最先端の理論までを説明する。6章は、これらの理論を計算機上に実装するための注意点についてまとめる。この章の目的は、論文で提案された理論(数式)をそのまま実装しても、実音声の分析において期待する性能が達成できるとは限らないため、実音声を計算機上で分析することを想定した細かな工夫を示すことである。7章では、推定された音声パラメータを加工する事例について紹介する。8章では、提案された音声処理技術の有効性を評価するために必要不可欠な主観評価について、基礎的なものを紹介することとした。

最後に、本書を執筆する機会を与えてくださった日本音響学会音響テクノロジーシリーズ編集委員会の飯田一博委員長、編集委員の北村達也氏をはじめとする委員会の皆様、本書の草稿に対し丁寧にコメントをくださった和歌山大学の河原英紀氏、大学入試センターの内田照久氏に深謝する。また、曜日を問わず自由気ままに執筆することを許してくれた妻に感謝する。


2018年5月

森勢 将雅


目 次

1. 基礎知識


1.1 本書で共通する数学的知識	2
1.1.1 音声波形とスペクトルについて	2
1.1.2 スペクトルの振幅, 位相	3
1.1.3 群 遅 延	5
1.1.4 信号の平均時間と持続時間	7
1.1.5 スペクトルを用いた平均時間と持続時間の表現	8
1.1.6 不確定性原理	12
1.1.7 畳 み 込 み	12
1.1.8 デジタルフィルタ	14
1.1.9 z 変換によるデジタルフィルタの特性解析	15
1.1.10 窓関数による波形の短時間分析とスペクトログラム	18
1.1.11 瞬 時 周 波 数	22
1.2 音 声 の 収 録	25
1.2.1 マイクロフォンによる音声の取り込み	25
1.2.2 近 接 効 果	27
1.2.3 収録環境の騒音レベル	28
1.2.4 収録環境の残響時間	32
1.2.5 信号対雑音比 (SNR)	33
1.3 A-D 変 換	34
1.3.1 標本化と量子化	35
1.3.2 折り返しひずみ	36
1.3.3 高品質な音声分析合成に求められる水準	38
引用・参考文献	40




2. 音声のパラメータ表現



2.1 音声の生成メカニズムと音声の分類	43
2.1.1 有声音の発声メカニズム	43
2.1.2 無声音の発声メカニズム	45
2.2 音声を構成するパラメータ	46
2.2.1 基本周波数	47
2.2.2 スペクトル包絡	49
2.2.3 非周期性指標	50
2.2.4 有声音の定式化	51
2.3 伝統的な音声分析合成システム	53
2.3.1 ボコーダ（非周期性指標を不使用）	53
2.3.2 ボコーダ（非周期性指標を使用）	62
2.3.3 フェーズボコーダ	64
2.3.4 正弦波モデル	68
2.4 本章のまとめ	71
引用・参考文献	71



3. 基本周波数の推定



3.1 古典的な方法	76
3.1.1 ゼロ交差法	76
3.1.2 自己相関法	78
3.1.3 ケプストラム法	80
3.1.4 共通する問題点	83
3.2 精度を高めるための工夫	84
3.2.1 相関法の改良	85
3.2.2 ゼロ交差法の改良	87

3.3 実用レベルにある最先端の方法	91
3.3.1 基本周波数候補の推定	92
3.3.2 基本周波数軌跡の推定	101
3.3.3 最終的な軌跡の確定と平滑化	105
3.4 基本周波数推定法の性能評価	108
3.4.1 Electroglottography (EGG) を用いた真値の定義	108
3.4.2 評価指標	110
3.4.3 基本周波数評価における課題	112
引用・参考文献	113



4 スペクトル包絡の推定

4.1 線形予測符号	116
4.1.1 問題設定	116
4.1.2 最適係数の導出	117
4.1.3 LPCの妥当性	120
4.1.4 LPCの問題点	121
4.2 ケプストラム	122
4.2.1 ケプストラムによるスペクトル包絡推定	122
4.2.2 ケプストラムの問題点	124
4.3 高品質音声分析合成のための課題の整理	124
4.4 STRAIGHT	125
4.4.1 STRAIGHT で用いる窓関数	126
4.4.2 平滑化とスペクトル補償	129
4.4.3 スペクトル補償の意味	132
4.5 分析時刻に非依存なスペクトル包絡推定の前提知識	133
4.5.1 スペクトルの離散化	134
4.5.2 窓関数により切り出す時刻の影響	135
4.5.3 分析時刻に依存した成分の定式化	136
4.6 TANDEM-STRAIGHT	138
4.6.1 TANDEM	139

4.6.2	平滑化とスペクトル補償	141
4.7	CheapTrick	142
4.7.1	分析時刻に依存した項の再解釈	143
4.7.2	窓関数の設計	143
4.7.3	パワースペクトルの変形	147
4.7.4	平滑化とスペクトル補償	147
4.8	スペクトル包絡推定法の性能評価	148
4.8.1	対数スペクトルに対するユークリッド距離	148
4.8.2	板倉・斎藤距離	149
4.8.3	周波数軸上で重み付けされた板倉・斎藤距離	150
4.8.4	スペクトル包絡推定評価の課題	151
	引用・参考文献	152

5. 非周期性指標の推定



5.1	前提となる考え方	156
5.1.1	雑音が重畳された音声の定義	156
5.1.2	HNR	156
5.1.3	非周期性指標推定の目標	157
5.2	STRAIGHT で用いる推定法	158
5.2.1	基本的な考え方	158
5.2.2	時間軸の非線形伸縮	159
5.2.3	非周期性指標推定	160
5.3	TANDEM-STRAIGHT で用いる推定法	164
5.3.1	基本的な考え方	164
5.3.2	非周期性指標推定	167
5.4	WORLD で用いる推定法	171
5.4.1	基本的な考え方	171
5.4.2	具体的な推定アルゴリズム	172
	引用・参考文献	182

 **6. 高精度に計算するコツ** 

6.1 窓関数による波形の厳密な切り出し	185
6.1.1 窓関数の厳密な設計	185
6.1.2 0 Hz 成分の厳密な除去	186
6.2 スペクトル包絡推定における 0 Hz 成分の扱い	188
6.2.1 0 Hz 成分が推定結果に与える影響	188
6.2.2 0 Hz 成分の制御	189
6.3 高精度なスペクトルフィルタリング	190
6.3.1 線形補間による矩形窓の畳み込み	191
6.3.2 リフタリングを用いた方法	191
6.4 1 サンプル未満の遅延の制御	192
6.4.1 サブハーモニックの影響	192
6.4.2 微細な遅延の付与	193
6.5 波形生成時における 0 Hz 成分の除去	195
6.5.1 0 Hz 成分の除去が知覚に与える影響	196
6.5.2 適切な 0 Hz 成分の除去	198
6.6 ボコーダにおける無声音の扱い	199
6.6.1 問題の設定	199
6.6.2 スペクトル包絡推定・波形生成における無声音の扱い	200
6.7 瞬時周波数計算における注意点	202
6.7.1 窓関数の影響	202
6.7.2 瞬時周波数計算における窓関数の差	203
6.7.3 発散する時刻の瞬時周波数	205
引用・参考文献	208

7. 音声の加工技術

7.1 基本周波数の加工	209
7.1.1 基本的な加工	209
7.1.2 抑揚の大きさの加工	210
7.1.3 基本周波数操作を行うための軸変換	212
7.2 スペクトル包絡の加工	213
7.2.1 加工に関する基本的な考え方	213
7.2.2 フィルタリングによるスペクトル包絡の加工	214
7.2.3 スペクトル包絡の伸縮による音色の加工	215
7.3 発話時間の加工	217
7.4 複数パラメータを組み合わせた加工	217
7.4.1 性別の変換	218
7.4.2 有声音のささやき声化	218
7.4.3 音高錯覚	219
7.5 音声モーフィング	221
7.5.1 時間・周波数軸上のラベル付け	221
7.5.2 時間・周波数軸の非線形伸縮	222
7.5.3 伸縮された時間周波数表現における加重平均	224
7.5.4 モーフィングの拡張	225
7.6 歌声合成への応用	226
7.6.1 歌声の高さに関する単位	226
7.6.2 微細構造	227
7.6.3 ビブラート	227
7.6.4 音高遷移に関する歌唱表現	231
7.6.5 歌唱フォルマント	234
引用・参考文献	235

 **8. 音声品質の主観評価方法** 

8.1 音声分析合成法に関する主観評価のおもな流れ	238
8.1.1 目的に基づく評価法の設計	238
8.1.2 実験規模の設計	239
8.1.3 実験結果の解析	239
8.2 共通する実験の前処理	240
8.2.1 音声の音圧レベルの正規化	241
8.2.2 実験刺激のランダム提示	242
8.2.3 実験環境の記録	243
8.3 リファレンスの有無にかかわらず利用できる評価法	244
8.3.1 MOS 評価	244
8.3.2 一対比較法	245
8.4 リファレンスに対する変化を測る評価法	249
8.4.1 DMOS 評価	249
8.4.2 CMOS 評価	250
引用・参考文献	251
索引	252

1 基礎知識

音声分析合成の学習を始めるためには、微分積分、線形代数などの数学や、デジタル信号処理に関する専門知識が必要不可欠である。その一方で、上記の領域のすべてをカバーしていることは、必ずしも必要ない。本書で扱う理論は一見複雑に見えるかもしれないが、丁寧にひも解くと、大学学部レベルで学ぶ知識の組合せである。ただし、どの部分をどの程度習得すればよいかの道標は必要と思われるため、本書で扱う理論を理解するために必要となる範囲の知識について、本章でまとめることとした。

音声分析合成の研究では、必要となる音声を収録することや、提案法の有効性を示すために音声の品質評価を行うことがある。本書では、このような読者が本書のみで問題を解決できるようにするため、前者については本章で、後者については8章でカバーすることとする。ただし、最低限の研究を行う入り口に立てることを目標として、深い議論を避け、必要に応じて論文や別の書籍へのリンクを付けることにした。本章で扱う内容は広く浅くなるため、詳細について興味を持ち深く理解したい読者は、是非とも関連書籍や論文に目を通していただきたい。

すでに、デジタル信号処理に関して一定の知識を有し、音声収録の経験もある読者は、本章をスキップしてもさしつかえない。もし、デジタル信号処理に関する基礎的な勉強が必要なら、文献1), 2) が入門編に適している。近年では学会がWebで知識を提供する試みもなされており、例えば文献3) は、デジタル信号処理に関する幅広い情報を提供している。より高度な内容を勉強する場合は、文献4) が参考になる。音声情報処理の入門的な内容については、文献5) が適している。

1.1 本書で共通する数学的知識

まず、さまざまな章で共通して使われる数学的な公式や理論、パラメータについて解説する。以下に示す記号や数式、公式などは、各章では証明なしで利用することとする。なお、理論を計算機上で実装するためには、おのおのの理論を離散信号を対象に記述する必要があるが、本書では、特に断りがない限り連続信号を対象に理論を導出する。離散系で導出したほうが容易となる場合は、そのことを示した上で離散信号を対象に導出する。

1.1.1 音声波形とスペクトルについて

まず、分析対象となる信号とフーリエ変換 (Fourier transform) により得られるスペクトル (spectrum) について整理する。入力信号 $x(t)$ 、スペクトル $X(\omega)$ の関係は、フーリエ変換、および逆フーリエ変換 (inverse Fourier transform) により以下の式で表される。

$$X(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x(t)e^{-i\omega t} dt \quad (1.1)$$

$$x(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} X(\omega)e^{i\omega t} d\omega \quad (1.2)$$

ここで、 t は連続時間を表し、 ω は連続的な角周波数を表す。 i は虚数単位であり $i^2 = -1$ である。虚数単位を j と表記する分野も存在するが、本書では i に統一する。フーリエ変換における積分記号の手前にある係数は、フーリエ変換、逆フーリエ変換のどちらか一方に $1/(2\pi)$ と記載することもある。本書では線形の理論のみを扱うため、適当な係数を乗じて正規化すれば、どの形でも結果に影響は生じない。係数の有無が計算を複雑にすることもあるため、状況に応じて上記のパターンを使い分ける。また、フーリエ変換は頻出するため、特にフーリエ変換の数式そのものが本質的に重要ではない場合には、それぞれ記号 $\mathcal{F}[\cdot]$ 、 $\mathcal{F}^{-1}[\cdot]$ を用いて以下のように簡略化する。

$$X(\omega) = \mathcal{F}[x(t)] \quad (1.3)$$

$$x(t) = \mathcal{F}^{-1}[X(\omega)] \quad (1.4)$$

離散表現に関しては、**離散時間**を表す記号を n 、**離散周波数**を表す記号を k と表記する。連続信号を離散信号に変換する処理については 1.3 節で概説するが、ここでは関連する知識を習得済みという前提で話を進める。**標本化周波数**が f_s の音声のとき、離散信号と連続信号との関係は、以下の式で表される。

$$x(n) = x\left(\frac{1}{f_s}n\right) \quad (1.5)$$

ここで、 n は整数であり、左辺は離散信号、右辺は連続信号である。離散周波数については、標本化周波数を f_s [Hz]、 $x(n)$ の点数を N 点とすると、以下の関係がある。

$$X(k) = X\left(\frac{f_s}{N}k\right) \quad (1.6)$$

N に制約を与えなければ**離散フーリエ変換** (discrete Fourier transform; **DFT**) であるが、ここでは、**高速フーリエ変換** (fast Fourier transform; **FFT**) を利用することを前提に、 N は 2 のべき乗であることとする。 k は整数であるが、**折り返しひずみ**の影響があり、値の範囲は 0 から $N/2$ までの $N/2 + 1$ 点まで記録しておけば、残りのスペクトルは計算できる。具体的に、 $N/2 + 2$ 点から $N - 1$ 点までは、以下のように 0 から $N/2$ の値の複素共役の関係にある。

$$X(N - k) = X^*(k) \quad (1.7)$$

ここで、 X^* は X の複素共役を表す。また、 $X(0)$ と $X(N/2)$ は実部のみ有することにも注意が必要である。 N 点の信号に FFT を施すと、0 点と $N/2$ 点の信号の虚部は 0 であり、1 点から $N/2 - 1$ 点の信号は一般的に複素数である。実部と虚部を独立した数値と考えると、FFT とは、 N 点の数列を同じく N 点の数列に変換することとなる。

1.1.2 スペクトルの振幅, 位相

音声分析により得られるスペクトル $X(\omega)$ は、特殊な条件を満たした場合を

除き、一般的に複素数の関数となる。そのため、スペクトル解析では振幅スペクトル $|X(\omega)|$ と位相スペクトル $\varphi(\omega)$ に分離することが多い。人間の聴覚は位相の変化よりも振幅の変化に敏感であるため、音声情報処理の理論の多くは振幅スペクトルやその2乗のパワースペクトルを対象としたものが多い。ただし、これを「位相は音色に影響しない」と解釈することは適切ではなく、影響は相対的に小さいだけであることを注意する。位相の変化が音色に影響を与えることは、さまざまな実験により示されている^{6), 7)†}。

振幅スペクトル、位相スペクトルは、 $X(\omega)$ から以下の式により与えられる。

$$X(\omega) = \Re[X(\omega)] + i\Im[X(\omega)] \quad (1.8)$$

$$|X(\omega)| = \sqrt{\Re[X(\omega)]^2 + \Im[X(\omega)]^2} \quad (1.9)$$

$$\varphi(\omega) = \angle X(\omega) \quad (1.10)$$

ここで、 $\Re[\cdot]$ と $\Im[\cdot]$ は、それぞれ実部と虚部を取り出す操作に対応する。振幅スペクトルを2乗したものはパワースペクトル $|X(\omega)|^2$ である。位相については、一部のプログラミング言語では atan2 などの関数名で計算できるが、厳密には以下の式により与えられる。

$$\varphi(\omega) = \begin{cases} \tan^{-1} \left(\frac{\Im[X(\omega)]}{\Re[X(\omega)]} \right) & \text{if } \Re[X(\omega)] > 0 \\ \tan^{-1} \left(\frac{\Im[X(\omega)]}{\Re[X(\omega)]} \right) + \pi & \text{if } \Re[X(\omega)] < 0 \text{ and } \Im[X(\omega)] \geq 0 \\ \tan^{-1} \left(\frac{\Im[X(\omega)]}{\Re[X(\omega)]} \right) - \pi & \text{if } \Re[X(\omega)] < 0 \text{ and } \Im[X(\omega)] < 0 \\ \frac{\pi}{2} & \text{if } \Re[X(\omega)] = 0 \text{ and } \Im[X(\omega)] > 0 \\ -\frac{\pi}{2} & \text{if } \Re[X(\omega)] = 0 \text{ and } \Im[X(\omega)] < 0 \\ \text{undefined} & \text{if } \Re[X(\omega)] = 0 \text{ and } \Im[X(\omega)] = 0 \end{cases} \quad (1.11)$$

本書では、後ほど位相スペクトルの周波数微分を計算する都合上、以下の式で

† 肩付き番号は章末の引用・参考文献を示す。

位相を定義することとする。上記の式では、 ω とは無関係な項が加算されることもあるが、これらの項は微分計算時にすべて0となるため、以下のみでよい。

$$\varphi(\omega) = \tan^{-1} \left(\frac{\Im[X(\omega)]}{\Re[X(\omega)]} \right) \tag{1.12}$$

1.1.3 群 遅 延

位相は、複素数を対象に計算される、複素平面上の角度に相当する。したがって、 2π の範囲でしか値は与えられず、例えばある周波数の位相が0だとしても、それが0なのか 2π なのかは区別できない。図 1.1 のように位相スペクトル

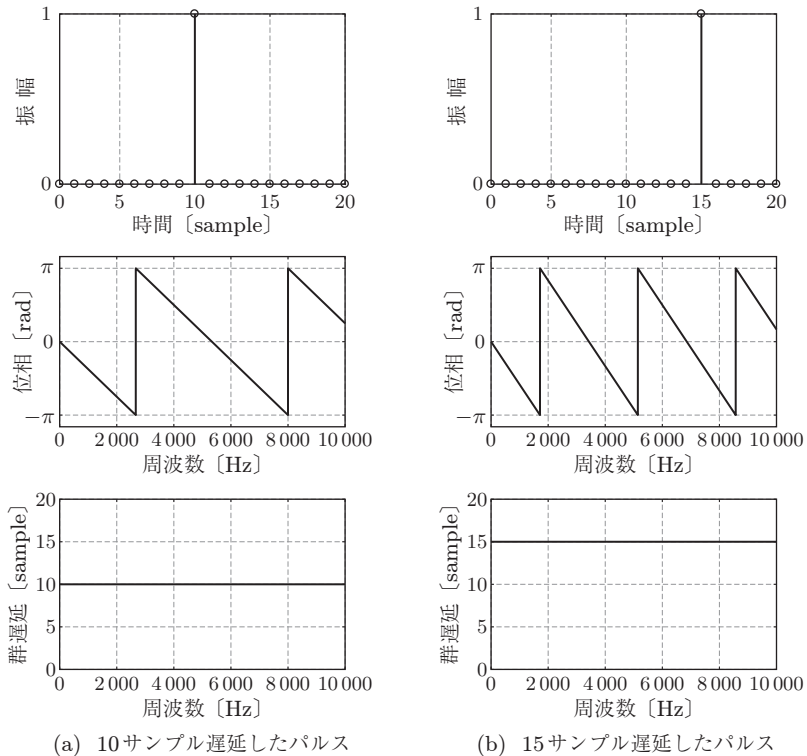


図 1.1 波形と位相と群遅延との関係。位相を $-\pi$ から π の範囲で定義しているため、不連続なジャンプが発生している。群遅延は、各周波数成分が遅延した時刻に相当するため、時間的にシフトしたパルスについて計算すると、全周波数で均一な時刻だけシフトしている。

索引

【あ】

圧力傾度型マイクロフォン 27

安定性 16

【い】

位相スペクトル 4

位相のアンラップ 6

板倉・斎藤距離 149

一対比較法 245

移動平均モデル 17

因果性 55

イントネーション 47, 210

インパルス応答 15

【う】

ウィーナー・ヒンチンの
定理 78

ウェーバー・フェヒナーの
法則 36

ウェルチ法 140

歌声モーフィング 225

【お】

オイラーの公式 66

オクターブバンド分析 31

オーバーシュート 231

折り返しひずみ 3

音 圧 28

音圧レベル 28

音響管モデル 55

音 高 47

音高錯覚 219

音声符号化 46

音声変換 50

音声モーフィング 221

音 素 42

音素境界 221

音素表記 42

音 程 47

音 名 227

【か】

カイザー窓 165

外 挿 170

ガウス窓 126

歌唱フォルマント 234

可聴域 28

可聴音 28

カットオフ周波数 18

過渡応答 106

加法定理 145

ガンマ関数 166

【き】

基本周期 47

基本周波数 47

基本波 76

基本波フィルタリング 76

基本波らしさ 89

逆位相 127

逆行列 169

逆フィルタ 130

逆フーリエ変換 2

逆向マスキング 196

吸音率 33

球面波 27

共 振 44

共 鳴 44

共鳴周波数 120

極 座 標 17

極座標 9

距離関数 148

近接効果 27

【く】

矩形窓 19

区分線形補間 222

クリッピング 40

群遅延 6

【け】

経時マスキング 196

傾 度 27

ケプストラム 56, 115

ケプストラム法 80

ケフレンシー 81

減衰域 18

【こ】

高域強調 45

高域通過フィルタ 18

高速フーリエ変換 3

喉 頭 109

国際音声記号 42

誤差関数 247

コンデンサマイクロフォン 25

【さ】

最小位相 55

最小位相応答 55

最小可聴音圧 28

最小 2 乗法 168

サイドローブ 22

錯 視 219

嘆 声 156

サブハーモニック 193

差分方程式 15

三角窓 125

残 響 32

残響曲線 33

残響時間 32

サンプリング定理 35

【し】

子 音 43

時間軸の非線形伸縮 159

時間周波数解析 19

時間周波数表現 19

時間伸縮 67

時間的に不変なパワー
スペクトル 128

時間引き延ばしパルス 33

時間分解能 22

シグモイド関数 161

指向性 26

自己回帰モデル 17

自己相関関数 78

自己相関法 78

持続時間 7

時変性 151

時变的 91

時変フィルタ 70

遮断周波数 18

周期信号 48

周期性 74

周期性成分 50

収 束 15

十二平均律 226

周波数特性 15

周波数分解能 22

周波数マスキング 196

周波数レンジ 25

順向マスキング 196

瞬時音圧 28

瞬時周波数 22

瞬時振幅 230

順序効果 242

準同型ボコーダ 54

初期位相 70

初期反射音 32

ショートポーズ 105

信号対雑音比 33

振 動 18

振 幅 21

振幅スペクトル 4

振幅包絡 230

信頼度 96

心理量 28

【す】

スティーヴンスのべき法則 36

ステップ関数 197

スプライン関数 242

スペクトル 2

スペクトル解析 4

スペクトル重心 213

スペクトル包絡 44, 49

スペクトル補償 130

スペクトログラム 19

【せ】

静 圧 28

性格印象 212

正弦波モデル 47, 68

声質変換 209

声 帯 42

声帯振動 43

声 道 44

声道断面積関数 214

声道長 120

声道長伸縮 219

正方向列 169

セグメンタル SNR 34

ゼロ位相 55

ゼロ位相フィルタ 106

ゼロ交差点 76

ゼロ交差法 76

ゼロ点 17

遷移域 18

全域通過フィルタ 58

線音源 32

全極モデル 17

線形時不変システム 14

線形予測係数 116

線形予測符号 115

線スペクトル対 115

【そ】

騒音計 28

騒音レベル 28

相互相関 85

相互相関関数 85

相対パワー 21

相補的時間窓 126

ソース・フィルタモデル 46

ゾーン 29

【た】

帯域通過フィルタ 18

帯域幅 12

帯域分割 164

対応点 221

対数スペクトル距離 148

ダイナミックマイクロ
フォン 25

ダイナミックレンジ 39

ダウンサンプリング 38, 165

多重音解析 75

畳み込み 12

単一指向性マイクロフォン 26

短時間フーリエ変換 19

短時間分析 19

単調増加性 225

【ち】		【は】		フォルマントトラッキング	122
チャンネルボコーダ	42	倍音	89	フォン	29
超音波	28	倍ピッチエラー	74	不確定性原理	12
調音フィルタ	44	波形接続法	116	複素正弦波	202
聴覚マスキング	196	パーセバルの定理	9	複素フーリエ級数	52
重畳加算法	68	バターワースフィルタ	106	複素平面	5
調波	70	発散	15	物理量	47
調波構造	74	バートレット窓	125	部分自己相関係数	115
調波複合音	112	ハニング窓	20	ブラックマン窓	20
直接音	32	パラメトリック推定	115	フラッター	227
直線位相	6	パルス符号変調	35	ブランシュレルの定理	9
直流成分	117, 185	パルス列	47	フーリエ変換	2
直交ミラーフィルタ	164	破裂音	43	フルバンド音声	39
【つ】		パワースペクトル	4	プレパレーション	231
通過域	18	反射	27	フレーム	22
【て】		反射音	32	フレームシフト幅	22
低域通過フィルタ	18	半波整流	131	フレーム長	22
デジタルフィルタ	14	半ピッチエラー	48, 79	【へ】	
ディラックのデルタ関数	51	【ひ】		平滑化	16
点音源	32	非可逆圧縮	36	閉管	120
電源ノイズ	94	非最小位相応答	57	平均時間	7
伝達関数	16	非周期性指標	50	平面波	27
【と】		非周期性成分	50	ベースライン	238
等価騒音レベル	30	ピッチ	47	ベッセル関数	166
同時マスキング	196	ピッチ推定	75	ヘッドセットマイクロ	
動的圧縮型ガンマチャープ		ピッチ同期重畳加算	62	フォン	27
フィルタバンク	151	ピッチ同期分析	127	【ほ】	
等ラウドネスレベル曲線	29	非負	129	母音	43
【な】		非負性	131	防音室	27
ナイキスト周波数	6	ビブラート	69, 227	放射特性	44
ナットール窓	20	標本化	35	ボコーダ	46
滑らかな半波整流	131	標本化周波数	3, 35	ポーラパターン	26
【ね】		標本化定理	35	ポルタメント	231
音色	4, 49	ヒルベルト変換	56	【ま】	
【の】		【ふ】		摩擦音	43
ノンパラメトリック推定	116	ファンタム電源	25	マスクー	196
		不安定	18	マスキー	196
		フェーズボコーダ	47, 64	マスキング効果	36
		フォルマント周波数	44	窓関数	19

窓関数法	165	有声無声判定	49	離散信号	2
【む】		ユークリッド距離	148	離散フーリエ変換	3
無響室	27	揺らぎ	106	梨状窩	121
無指向性マイクロフォン	26	ユール・ウォーカー方程式	118	理想フィルタ	18
無声音	43			リファレンス	238
		【よ】		リフタ	124
		抑揚	210	リフタリング	124
				量子化	35
		【ら】		量子化ビット数	36
メインローブ	21	ラウドネスレベル	29	【る】	
メインローブ幅	22	乱数	157	累積分布関数	247
メディアンフィルタ	100			累積和	6
メル尺度	209	【り】		【れ】	
面音源	32	離散化	35	励起信号	53
		離散時間	3	連続信号	2
有意差検定	245	離散周波数	3		
有声音	43				



【A】		【D】		【H】	
A 特性	30, 242	D-A 変換	35	Harvest	92
A 特性音圧レベル	28	DC 成分	185	HNR	156
A 特性時間重み付き サウンドレベル	30	DFT	3	HPF	18
A-D 変換	34	DIO	88	【I】	
AR モデル	17	DMOS	249	IIR フィルタ	14
ARMA モデル	17	D4C	171	【L】	
ARX モデル	122	【E】		LPC	54, 115
		EGG	108	LPC ボコーダ	53
【B】		Electroglottography	108	LPF	18
BAP	155	ERB _N	150	LSP	115
BPF	18	【F】		【M】	
BT 積	12	FFT	3	M 系列信号	33
buzzy	50	FIR フィルタ	14	MA モデル	17
		FPE	110	mel	75
【C】		【G】		Mixed-source model	50, 155
CD 音質	39	GPE	110	MOS	148, 244
cent	226	Gross error	110	MRI	214
CheapTrick	142				
CMOS	249				

Multiband excitation							
vocoder	50, 155						
	[N]		[R]			[W]	
		RMS	110		WORLD		142
NC	30					[Y]	
	[O]	sinc 関数	38		YIN		85
		SPL	28			[Z]	
OLA	68	STFT	19		z 変換		15
	[P]	STRAIGHT	116		Zero-lag butterworth filtering		108
PARCOR	115				~~~~~		
PESQ	148, 237	[T]				[数字]	
POLQA	148, 237	TANDEM	126				
PSOLA	62	TANDEM-STRAIGHT	126				
	[Q]						
		[V]					
QMF	164	VAD	91		2 乗平均平方根		110
		VTAF	214				

—— 著者略歴 ——

森勢 将雅 (もりせ まさのり)

2002年 釧路工業高等専門学校情報工学科卒業

2004年 和歌山大学システム工学部デザイン情報学科卒業

2006年 和歌山大学大学院システム工学研究科博士前期課程修了

2008年 和歌山大学大学院システム工学研究科博士後期課程修了 (短期修了), 博士 (工学)

2008年 関西学院大学博士研究員

2009年 立命館大学助教

2013年 山梨大学特任助教

2017年 山梨大学准教授

現在に至る

音声分析合成

Speech Analysis and Synthesis

© 一般社団法人 日本音響学会 2018

2018年8月6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 牧製本印刷株式会社

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 978-4-339-01137-1 C3355 Printed in Japan

(新井) G



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。