

大容量化する

マルチメディア・データを転送・保存・活用するために

デジタル音声&画像の 圧縮/伸張/加工技術

見本

●尾知 博 監修 川村 新, 黒崎正行 著

ノイズの除去や
複数の音声を分
離する方法も



第1章

信号処理で重要となる 基礎知識

1-1 確率信号の扱い方

音声信号は、マイクロホンなどにより、電圧や電流の振幅変化として表現できます。

音声信号において、ある時刻における振幅値は、実際に発話してからでなければ知ることができません。つまり、未来の値を確定することはできません。しかし、長時間かけて音声信号の値を観測すると、何らかの統計的性質が見えてきます。従って、音声信号を、ある統計的性質に従って生じる確率過程 (Stochastic Process) として扱うことにより、過去やある程度未来の信号の統計量は特定可能です。

ここでは、音声や画像の処理システムを構築する際に必要となる、確率信号の取り扱い方について説明します。

1.1.1 確定信号と確率信号

信号には、図1.1に示すように、確定信号 (Deterministic Signal) と確率信号 (Stochastic Signal) があります。

確定信号は、時刻や位置の関数として表される信号です。従って、時刻、あるいは位置だけを指定すると、特定の数値として表現できます。

確率信号は、確率過程として生じる信号です。時刻や位置だけの関数として表現することはでき

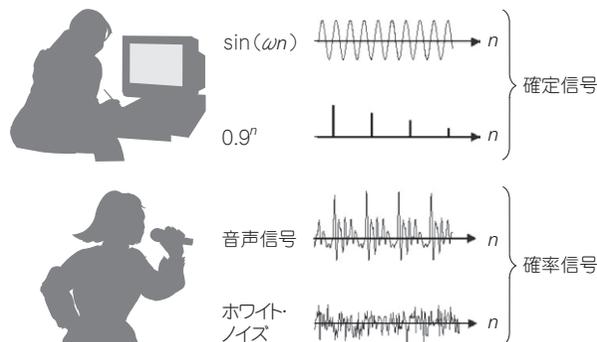


図1.1 確定信号と確率信号

第2章

音声圧縮技術

携帯電話や小型の音楽プレーヤの普及により、複数の音声を同時かつ高速に伝送することや、小型機器にさらに大量の音楽を保存することが望まれています。そのためには、音声や音楽をできるだけ小さいデータ量で表現しなければなりません。このような目的を達成するための技術を音声圧縮 (Audio Compression) といいます。現在、表2.1に示すような音声圧縮技術が用いられています。本章では、固定電話やコンピュータのwavファイルに用いられる方式と、携帯電話やVoIP^{注1}に用いられる方式について説明します。

音声圧縮が行われた後のデータを符号 (Code) といいます。また、符号から元の音声波形に戻すことを復号 (Decode) といいます。さらに、それぞれの操作を符号化、復号化といいます。

まず、音声波形そのものを符号化の対象とした波形符号化方式について述べ、次いで音声の発生機構を利用した分析合成方式について説明します。そして最後に聴覚特性を考慮して圧縮効率を高める方式について説明します。説明の一部として、Scilab^{注2}を用いた演習も行います。

表2.1 音声圧縮技術の分類

圧縮方式	用途	関連機関
PCM	CD, DVD, BD, ほか	ITU-T (G.701)
log-PCM	固定電話	ITU-T (G.711)
ADPCM	PHS, スーパーファミコン	ITU-T (G.726)
CELP	携帯電話, VoIP (Speex)	ITU-T (G.728, 729)
ATRAC	MD, ソニー製品各種, 音楽配信	ソニー
MP3	携帯音楽プレーヤ, インターネット・ラジオ, 音楽配信	MPEG
AAC	地デジ放送, iTunes, iPod, ほか	MPEG
ALS	スタジオ編集, 音楽配信	MPEG
AC-3	映画, DVD, BD, プレイステーション3, ほか	MPEG
DTS	映画館, DVD, BD, プレイステーション3, ほか	Dolby Laboratories社
RealAudio	RealPlayer, 音楽配信, デジタル録音	RealNetworks社
WMA	Windows Media Player, デジタル録音	Microsoft社

注1: Voice over Internet Protocol. インターネット上で音声データを送受信する技術。インターネット電話などに利用される。

注2: Scilabの導入と使用方法については本書のwebページにある「Scilabのインストール・操作方法」を参照。

第3章

ノイズ除去技術

第2章において説明した音声圧縮は、音声波形そのものをできるだけ小さい容量で表現するための技術でした。ここでは、音声波形の保存ではなく、音声とノイズが混在する波形を加工して音声波形だけを取り出す技術について説明します。

3-1 適応ノッチ・フィルタで正弦波を除去してみよう

回転体から生じるノイズやハウリング^{注1}は、極めて狭い周波数帯域に大きなパワーを持ちます。このようなノイズを狭帯域ノイズ (Narrow-band Noise) と呼びます。狭帯域ノイズを除去するためには、ある特定の周波数だけを除去し、ほかの周波数はすべて通過させるようなフィルタが有効です。このような特性を保持したまま、自動的に狭帯域ノイズを検出して除去する適応フィルタとして、適応ノッチ・フィルタ (Adaptive Notch Filter) があります。

適応ノッチ・フィルタによるノイズ除去のイメージを図3.1に示します。また、図3.2に適応ノッチ・フィルタの応用例を示します。

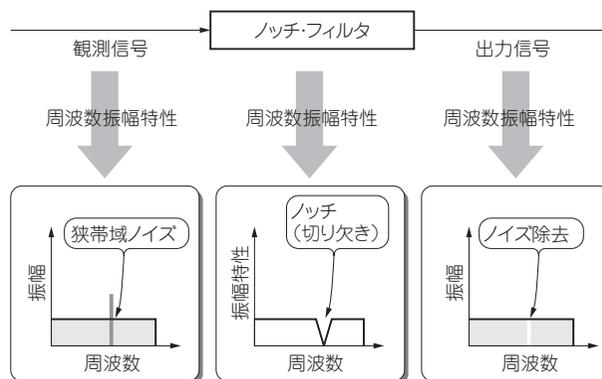


図3.1 適応ノッチ・フィルタによるノイズ除去

注1：スピーカとマイクロホン間に閉ループが形成されたとき、ある条件において単一の周波数振幅が急激に増大する共振現象のこと。カラオケなどでしばしば生じる。

第4章

音源分離技術

前章で説明した単一マイクロホンによるノイズ除去は、発話者が1名でかつノイズの特性がほぼ変化しない状況を想定していました。従って、特性変動の激しい音声が複数混在する場合に、これらを分離する目的には使用できません。

本章では、マイクロホンを複数に増やし、複数の音源を分離する技術について説明します。各音源から放射された音は、それぞれ異なる経路を通過してマイクロホンに到達します。各音源から各マイクロホンまでのそれぞれの経路の特性を総じて混合過程といいます。通常、混合過程は未知ですから、マイクロホンの観測信号だけを頼りに音源を分離しなければなりません。このように、混合過程が未知の観測信号から各音源を分離する方法を、ブラインド音源分離 (BSS: Blind Source Separation) と呼んでいます (図4.1)。

4-1 右と左に音源あり！バイナリ・マスクで音の持ち主を見分ける

最初に最も単純なBSS法である、バイナリ・マスクングについて説明します。図4.2に示すように、二つのマイクロホンに2人の話者が発話する場合を考えます。ここで、話者Aと話者Bは、図4.2の中央の破線の左側と右側にそれぞれ位置するものとします。

話者Aの音声は、まずマイクロホン1に到来し、次いでマイクロホン2に到達します。このとき、マイクロホン2で観測される音声は、距離が遠い分、マイクロホン1よりも振幅が減衰します。逆に、マイクロホン1で受信される話者Bの音声は、マイクロホン2よりも振幅が減衰します。もし、マ

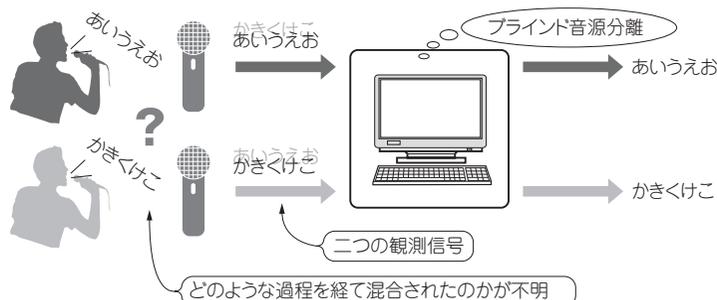


図4.1 ブラインド音源分離

第5章

映像メディア処理で重要となる基礎知識

現在、画像や動画などさまざまな場面において多次元の信号を扱うことが必要不可欠となってきました。本章では、この信号を拡張した多次元信号を扱い、多次元システムの基本について説明すると共に、2次元信号である画像を用いて多次元信号 (Multidimensional Signal) の信号処理について述べます。また、圧縮技術において必要となる直交変換についても例を用いて説明します。

5-1 画像やメディア処理で使われる信号

5.1.1 多次元信号

最初に、本章で扱う信号やシステムについて定義します。ここでは、複数の次元を有する信号を扱っていきます。これらの複数の次元を持つ信号を多次元信号といいます。同様に、複数の次元を有するシステムを多次元システム (Multidimensional System) といいます。

時間により音の強弱が変化するような一つの情報に対して変化を行う信号は、1次元信号 (One-dimensional Signal) です。静止画のように、 x 方向、 y 方向に信号が広がっている信号は2次元信号 (Two-dimensional Signal) です。静止画が時間 t ごとに变化していく動画信号は3次元信号 (Three-dimensional Signal) です。図5.1に1～3次元の信号の例を示します。

これらの信号はそれぞれ、式(5.1)～式(5.3)で表すことができます。

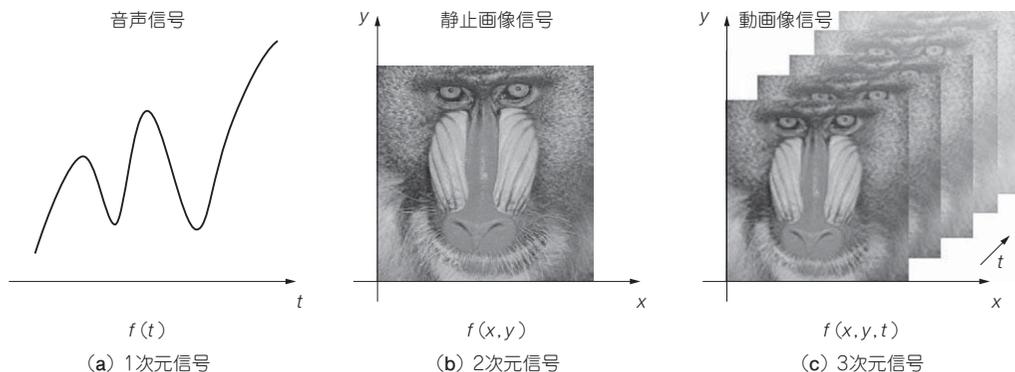


図5.1 N 次元信号の例

第6章

静止画像を圧縮するJPEG

フィルムのカメラからデジタル・カメラ、ビデオ・カメラからデジタル・ビデオ・カメラ、ビデオからBlu-rayディスクなど、さまざまな機器がアナログからデジタルへと変化してきています。しかし、アナログ信号を直接デジタル信号へ変換し蓄えると、膨大な量のハード・ディスクやメモリが必要となってしまいます。

例えば、ハイビジョン・テレビ (High Definition Television) を直接ハード・ディスクに蓄える場合を考えます。画像サイズは1920×1080画素であり、1画素当たり24ビット^{注1}、1秒間に30フレーム^{注2}とすると、1秒間で約1.5Gビット (約190Mバイト) のデータ量になります。2時間の番組であれば10.8Tビット (1.4Tバイト) です。

そのため、これらのデータを蓄積したり、伝送したりする場合、データを圧縮する必要があります。しかし、一般的なデータ圧縮では、データを元に戻すことはできるものの、高い圧縮率で画像圧縮することは難しくなります。

画像などの映像メディアでは、人間の見た目において問題ない画質であれば圧縮前の画像のデータと一致する必要がありません。このような考えに基づき映像メディアでは、高い圧縮率を達成しています。本章では、静止画像の画像圧縮について説明します。

静止画像の画像圧縮では、現在デジタル・スチル・カメラで利用されているJPEGについて説明します。JPEGよりも高画質に画像を圧縮できる静止画像の圧縮規格であるJPEG 2000については第7章で説明します。動画は複数の静止画像を連ねたものとして構成されています。この1枚1枚を静止画像として画像圧縮することで、動画圧縮も可能であるため、静止画像の画像圧縮は動画の画像圧縮にも利用されています。

6-1 画像や人間の特性を利用して圧縮するJPEGの概要

JPEGとは、Joint Photographics Expert Groupの頭文字をとったものであり、もともとは、静止画像の符号化を行うグループの名前でした、静止画像の圧縮規格の名称として定着しています。

JPEGは、国際標準化機構 (International Organization for Standardization : ISO)、国際電気標準会議 (International Electrotechnical Commission : IEC) と、電気通信に関する国際標準化機関

注1 : RGBそれぞれ8ビットずつとする。

注2 : 地上デジタル放送では、1秒間で29.97フレームであるが簡単化のため1秒間30フレームとして計算する。

第7章

映画館で利用されている JPEG 2000

デジタル画像の増大やコンピュータの性能の向上により、通信のみならずテレビや映画において、デジタル化が進んでいます。そのため、従来のJPEGより同じ圧縮率でより画質の高い画像圧縮規格が必要となってきました。そこで、誤り耐性機能や各種スケーラビリティ機能など、新たな機能を持ち、より高性能な圧縮可能な符号化画像の国際標準化が行われ、ISO/IECにおいて、JPEG 2000符号化は規格化されました。

JPEG 2000では、圧縮しても元の画像に戻すことのできる可逆符号化(Reversible Coding)、元に戻らないが、圧縮率を上げることのできる非可逆符号化(Inreversible Coding)を同じ構造で処理できるという特徴があります。

これらはJPEGと同じように、国際標準化機構(ISO)、国際電気標準会議(IEC)と、ITU-T、の共同作業において標準化された規格です。そのため、JPEG 2000はISO-15444-1/ITU-T Rec.800と二つの規格番号を持ちます⁽³⁵⁾。

一方、日本においても、日本工業規格(JIS)においてJIS X4350-1:2002、「JPEG 2000画像符号化システム—第1部：基本符号処理」という名称でJIS規格化されています⁽³⁶⁾。

加えて、動画像符号化においても、JPEG 2000符号化をフレームごとに施すMotionJPEG 2000符号化もISO-15444-3/ITU-T Rec.802やJIS X4350-3として標準化されています⁽³⁷⁾。この章ではJPEG 2000符号化の概要と誤り耐性機能について説明します。

7-1 JPEG 2000符号化の概要

JPEG 2000符号画像は図7.1により生成されます。JPEG 2000復号器の構成を図7.2に、JPEG 2000符号化処理の例を図7.3に示します。

JPEG 2000符号化は、JPEGと同様に直交変換による周波数成分への変換、量子化(必要に応じて処理を行う)、エントロピー符号化という同様の処理手順ですが、使用されるものが異なります。

入力される画像信号は、一つの色成分が8ビットの場合0～255となり、画像データの平均は正の値となります。これを直流成分といいます。この直流成分を取り除くことで、直交変換のための0平均のデータを作成します。その後、JPEGの色変換と同様の処理であるコンポーネント変換(Component Transform)が行われます。

次にJPEGでは、周波数成分の解像度は一定であるDCTを用いていましたが、JPEG 2000では、

第8章

動画を圧縮する MPEG/H.264

MPEGとは、Moving Picture Expert Groupの頭文字をとったものです。もともとは、動画像の符号化を行うグループの名前でしたが、現在は、動画像の圧縮規格の名称として定着しています。

MPEG符号化は、現在、動画像を圧縮する規格として広く普及しており、DVDビデオやデジタル放送、移動体通信で利用されています。特に、2014年から4K放送、2016年から超高精細テレビ規格である8Kの試験放送が始まるなど、高解像度の動画像を圧縮する必要が出てきています。それに合わせて、MPEG符号化は扱う対象に応じてMPEG-1からMPEG-21やMPEG-Aから-Vまでを国際標準として規格化および国際標準化中です。また、ITUもH.261をはじめとしてH.264の国際標準化を行っています。現在、ITUやMPEGでは、H.264を超える規格として、H.265やHEVC (High Efficiency Video Coding) を規格化しています。本章では動画を圧縮する基本として、MPEG-1、MPEG-2、MPEG-4 AVC/H.264について説明します。MPEG-1、MPEG-2、MPEG-4の応用について表8.1に示します。

8-1 動画を圧縮するMPEG符号化の仕組み

MPEG符号化画像の生成は、図8.1の符号化器を用いて行われます。MPEG符号化は、画像内の冗長性と時間の冗長性を削減するために、それぞれ、DCTと動き補償 (Motion Compensation : MC) を基本要素として構成しています。

自然画像では、フレーム内の注目している画素の近くは似ている場合が多く、画素の値の変化が小さくなります。そのため、空間的な周波数分解を行うと、周波数の低い領域に値が集中します。

表8.1 MPEG符号化の概要

符号化	伝送速度 [bps]	応用
MPEG-1 ビデオ	1M ~ 1.5M	蓄積メディア、CD-ROM
MPEG-2 ビデオ	5M ~ 100M	DVDビデオ、デジタル放送
MPEG-4 ビジュアル	20k ~ 20M	インターネット、CG、無線通信、スタジオ
MPEG-4 AVC/H.264	10k ~ 240M	次世代デジタル放送、Blu-ray ビデオ

見本

このPDFは、CQ出版社発売の「デジタル音声&画像の圧縮/伸張/加工技術」の一部見本です。

内容・購入方法などにつきましては以下のホームページをご覧ください。

内容 <http://shop.cqpub.co.jp/hanbai/books/31/31451.htm>

購入方法 <http://www.cqpub.co.jp/order.htm>



デジタル信号処理シリーズ