❶ 音楽と信号処理

基般

亀岡弘和（NTT）

音楽の信号処理

携帯音楽プレイヤや音楽配信サービスの普及，データストレージの大容量化などに伴い，楽曲やアーティストの検索，新しいスタイルの音楽鑑賞方法，楽曲提供者の著作権保護などをサポートする技術が重要になっている．音楽音響信号から音楽的に意味のある情報を自動的に取り出す音楽信号処理技術はこれらを実現する上で不可欠である．

音楽は，人間が発し聴く音のメディアとして音声と双璧をなしており，音楽信号処理と音声信号処理の研究は関連が深い一方で異なる面も多い．本特集の記事「3. 音楽と音声情報処理」でも触れられているように音楽と音声には主に 4 つの相違点が挙げられる．まず第 1 に，音声においては音韻が言語的な役割を担っているのに対し，音楽においては旋律，リズム，和声がその役割を担っている点である．その意味で，音楽から音高，リズム，和音を認識するのは音声における音声認識に相当している．第 2 に，音声とは異なり音楽ではほとんどの場合複数の音が混在していることが前提になっている点である．通常，音声信号処理や音声認識では対象となる音声は 1 つであり，それ以外の音（雑音）の影響をいかに回避するかなどが課題となるが，音楽では対象そのものが複数の楽音からなる．後述する多重音解析と調波打楽器音分離は，多重音から各楽音の基本周波数（音高）や打楽器音成分を推定する技術である．第 3 に，音楽はリズムという強い時間的秩序を有している点である．リズム・ビート解析は音楽音響信号からリズム・ビートを推定するための技術である．第 4 に，音楽は大域的な繰り返し構造や共通構造を有している点である．たとえば，ポピュラー音楽では A メロやサビといったセクションが楽曲中に繰り返される．楽曲構造解析はこのような大域的構造を捉えるための技術である．次章で，音楽信号処理の重要トピックを紹介する．

音楽信号処理の主なタスクと手法

♫ 多重音解析・音源分離

ヴァイオリンなどのようにピッチ（音の高さ）のある楽音の信号は局所的に周期的である．周期信号を構成する周波数成分の中で最も低い周波数を基本周波数という．多重音解析とは，複数の楽音が重畳した混合信号から個々の楽音の基本周波数（F_0）を推定する問題である．音楽音響信号の基本周波数は曲を特徴づける最も重要な情報の 1 つでこれを自動獲得できれば自動採譜，楽音分離，音楽検索などさまざまな応用に有用である．音声信号処理の分野でも基本周波数推定の研究は長く行われてきたが，そのほとんどは単一音が対象であった．

多重音解析の問題は音源分離の問題と密接に関係している．これを示すため，まず単一音のスペクトルから基本周波数を推定する問題を考える．信号が純音の場合，スペクトルのピーク周波数が基本周波数に対応する（**図 -1 (a)**）が，一般の周期信号には調波成分に対応する複数のピークがある（**図 -1 (b)**）．そして複数あるピークのうち最大のピークの周波数が必ずしも基本周波数に対応するとは限らない（**図 -1 (c)**）．また，基本周波数成分はい

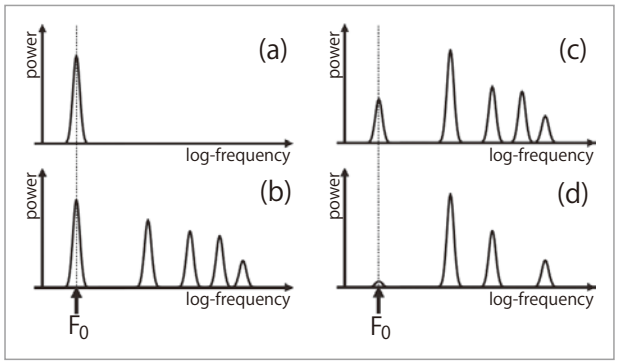


図 -1　基本周波数推定の問題

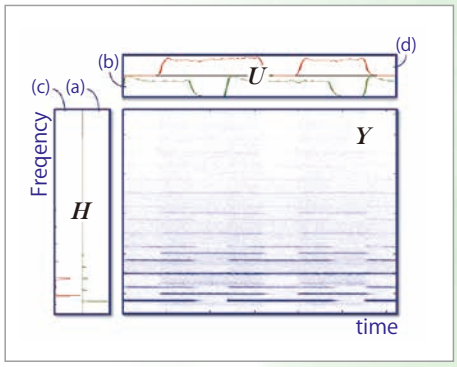


図 -2
NMF によるスペクトログラムの分解

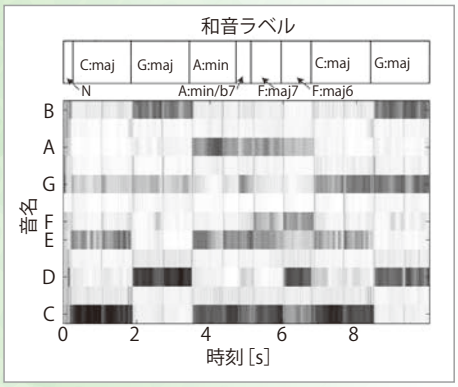
つも大きいとは限らないため，複数あるピーク周波数のうち最も低い周波数が必ずしも基本周波数に対応しない（**図 -1（d）**）．したがって，基本周波数を推定するためには対象とする音の信号波形やスペクトル構造の全体を手がかりにした方法が必要になる．しかし，多重音の音響信号には，各周波数でどの程度の成分がどの音に帰属するのかという情報が欠落しているため，基本周波数を推定するための重要な手がかりが得られないのである．

各音源の波形やスペクトルに関する先験知識が得られる場合には，基本周波数をパラメータに持つパラメトリックモデルを用いて観測信号または観測スペクトルにフィッティングする手法が有効である．一方，各音源のスペクトル構造に関する詳細な仮定を置く代わりに，各音源のスペクトルが観測区間において繰り返し生起するという仮定に基づく音源分離手法が提案されており，近年強力なアプローチとして注目されている．たとえば**図 -2** の（a），（c）のようなスペクトルの音が（b），（d）のような音量軌跡で鳴っていたとする．スペクトログラムが加法的であれば，これら 2 種類の音の多重音のスペクトログラムは，（a）と（c）を横に並べた行列 $\boldsymbol{H}$ と（b）と（d）を縦に並べた行列 $\boldsymbol{U}$ の積によって表される．これは逆に，観測された多重音のスペクトログラムを行列と見なし，2 つの行列の積に分解することにより各音源のスペクトルおよび音量軌跡の情報が得られることを意味する．ただしスペクトルは非負値なので，各行列の要素が非負となるような制約が必要であることから，このアプローチは非負値行列因子分解（Non-negative Matrix Factorization : NMF）と呼ばれる．

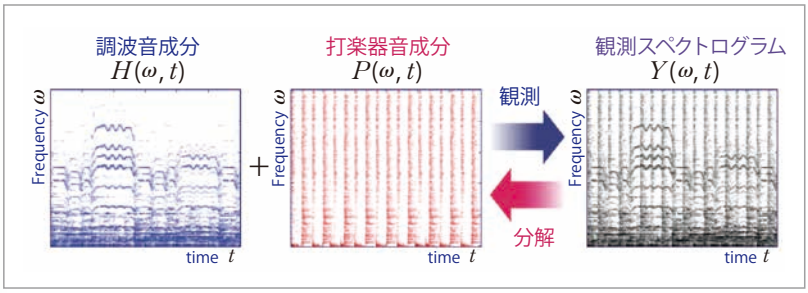
♫ 調・和音推定

西洋音楽やポピュラー音楽などにおいて調や和音は旋律やリズムと並ぶ楽曲の重要な構成要素である．音楽音響信号から各時刻での調・和音を推定する問題をそれぞれ調推定・和音推定と呼ぶ．通常，調や和音が同一の区間においても各時刻では構成音の音高は多様に変化するため，各時刻周辺の観測信号のみから調や和音を一意に決定することはできない．また通常，調や和音が変化するタイミングは未知である．したがって，調・和音推定では調・和音区間推定と各区間における調・和音同定の問題を解く必要がある．もし音楽音響信号中で調や和音が同一の区間が分かれば，当該区間において出現する音高の頻度などを手がかりに調や和音を推定することができる．一方，調や和音の出現順序が既知であれば調や和音が変化する時刻を推定することが可能である．このように，調・和音区間推定と各区間における調・和音同定の問題は相互依存の関係にある．

以上の性質の問題のため，隠れマルコフモデルやその拡張モデルを用い，同一和音（または調）の区間推定と各区間の和音（または調）推定の同時解決を目指した手法が有効である．和音特徴量として，スペクトログラムを音名ごとにオクターブ間で足しあわせたクロマグラムがしばしば用いられる．**図 -3** を見ると，同一の和音において特徴量が類似していることが確認できる．



◀図 -3　クロマグラム



▲図 -4　調波打楽器音分離の問題

♫ 調波打楽器音分離

クラシック音楽やポピュラー音楽ではピッチのある楽音（以後，調波音）と打楽器音が混在することが多い．前者には主に旋律や和声を表現する役割があるのに対し，後者には主にリズムを表現する役割がある．多重音解析と和音認識では音楽音響信号の中の旋律や和声，リズム解析やビート解析ではリズムに関する情報を抽出することが目的であるため，音楽音響信号をこれらの 2 つのタイプの音に分離する技術が有用となる場面は多い．また，調波音と打楽器音を分離できれば，それぞれの音量を変更できる音楽再生システムを提供することもできる．これを実現する技術を調波打楽器音分離という．しかし，たとえば 7+3 を解くのは簡単でも X+Y=10 となる X と Y を一意に決められないのと同様で，一般に一度混ざり合った信号を分離する問題は難しい．

図 -4 のとおり，調波音は周波数成分が時間方向に平行に連なる一方で，打楽器音は周波数方向に平行に連なる傾向にある．前者は，同一音高が一定時間持続することにより各調波音の調波構造中のピークが時間方向に平行に連なることによる．一方後者は，広帯域におよぶスペクトルが打叩時に急峻に立ち上がりすぐに減衰するためである．筆者らは，調波音と打楽器音においてスペクトログラムに現れるこれらの傾向に着目し，画像処理的なアイディアにより観測スペクトログラムを調波音と打楽器音の成分に分解する方法を提案し，Harmonic / Percussive Signal Separation（HPSS）法と呼んでいる．これ以外のアプローチとして，前述の NMF に基づくアプローチなども提案されている．

♫ ビート解析

音楽にはほぼ等間隔に繰り返される基本的なリズムがある．これを拍（ビート）といい，音楽音響信号や MIDI（Musical Instrument Digital Interface）信号から各拍の時刻や拍の間隔（テンポ）を推定する問題をそれぞれビート解析，テンポ解析という．実際の演奏において，拍は必ずしも正確に等間隔に打たれるわけではなく，演奏の表情付けなどによりその間隔は揺らぐことが多い．また，すべての拍位置で音が発せられるとは限らないし，拍位置以外で音が発せられることもあるため，音のありなしの情報だけではビートやテンポを推定することはできない．

拍はほぼ等間隔であること，拍位置において和音が変わりやすいこと，各音が拍位置で発せられる可能性が高いこと，などが本問題の解決の手がかりとなる．そこで，各時刻において発音された音が存在した確率を表すオンセット特徴量の系列から，隠れた周期的なピークを捉えるアプローチが有効である．オンセット特徴量としてはスペクトル変動量や深層学習により得られる特徴量，特徴量系列の周期性を捉える方法としては短時間フーリエ変換，隠れマルコフモデル，動的計画法を用いた手法などが提案されている．

♫ 楽曲構造解析

楽曲構造解析とは，音楽音響信号をセグメントに分割し，各セグメントを何らかのカテゴリ（ポピュ

ラー音楽のサビや A メロ，ソナタ形式の楽曲の提示部や展開部）に分類する問題である．この技術はサビの自動検出や楽曲のサムネイル（試聴用音源など）自動生成などさまざまなアプリケーションに役立つ．構造を基礎づける音楽の構成要素の関係性は「新規性」，「同質性」，「繰り返し構造」といった基準によって作られる．たとえば，新規のセクションの開始時にはフィルインなど突然の変化が生じる傾向にあり，同一のセクションの区間では調やテンポ，楽器編成などが一貫している傾向にある．また，ポピュラー音楽の 1 番と 2 番のサビなどのように，旋律や和音系列，リズムパターンなどが繰り返し用いられていれば同一のセクションと見なせる．これらを手がかりに，楽曲全体に隠れた構造をいかにして見い出すかが本問題の課題となる．

図 -5 は，二時刻間の特徴量の類似度を各要素にした自己類似度行列を示している．自己類似度行列上で，近辺の類似度が高いブロック状の個所（右上部の実線で囲まれた部分）は同質性が高く，対角上にあるブロック同士の継ぎ目（左下の「＋」が指し示す点）で新規性が高い．非対角成分上で対角に走る線が繰り返し構造を表している．新規性に着目したアプローチではブロック同士の継ぎ目を見つけ出す問題として定式化され，変化点検知に基づく手法が提案されている．同質性に着目したアプローチではセグメントをクラスタリングする方法や，非対角成分上の対角に走る線を動的計画法や画像処理の手法を用いて検出する方法が提案されている．

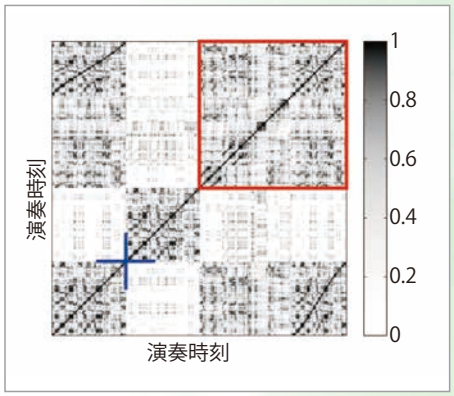


図 -5
自己類似度行列の例

音楽信号処理のこれから

本稿では，多重音解析，調・和音認識，調波打楽器音分離，ビート検出，楽曲構造解析など，音楽音響信号処理における重要課題と手法を紹介した．音楽信号処理研究の黎明期は音声の分野で長く培われた方法論や技術を導入しようという事例が多く見られたが，冒頭で述べたように音楽には音声にないさまざまな固有の特徴があることから，音楽ならではの独自の信号処理技術が近年発展してきている．特に音楽信号処理では多重音を扱うことが必須であることから NMF をはじめとした音源分離の研究が非常に進んでおり，音声信号処理（雑音・残響除去）の分野でも注目されている．一方で，最近は音声分野と足並みをそろえるかのように深層学習を各種タスクに適用する研究が盛んに進められているが，音声分野と比べて研究コミュニティがまだまだ小さいこともあり学習データセットを効率的に構築できる環境が整っているとは言えない．今後こうした環境が整備され，深層学習によるブレイクスルーが音楽信号処理分野でも起これば，まだまだ解決すべき課題の多い自動採譜，音楽検索・推薦の問題に突破口が見つかる可能性がある．

（2016 年 4 月 1 日受付）

亀岡弘和（正会員） kameoka.hirokazu@lab.ntt.co.jp

2002 年東大・工・計数卒業．2007 年同大学院博士課程修了．同年日本電信電話（株）入社．NTT コミュニケーション科学基礎研究所配属．2011 年東大大学院情報理工学系研究科客員准教授．2016 年国立情報学研究所客員准教授．音声・音楽を対象とした音響信号処理・機械学習の研究に従事．日本音響学会，電子情報通信学会，IEEE 各会員．情報理工学博士．IEEE Signal Processing Society 2008 SPS Young Author Best Paper Award 等受賞多数．