音楽音響信号処理技術の最先端

Recent Advances in Music Signal Processing Techniques

亀岡弘和　中村友彦　高宗典玄

Abstract

インターネットと携帯音楽プレーヤの普及に伴い，音楽検索システムや音楽鑑賞インタフェースなど，音楽に関連した新しいシステムやインタフェースの研究の重要性が高まっている．音楽音響信号処理技術は，計算機で音楽を扱うあらゆる応用システムの基礎となり得，今や世界中で研究が行われている．本稿では，多重音解析，調・和音認識，リズム解析，調波打楽器音分離，ビート解析，楽譜追跡，楽曲構造解析など，音楽音響信号処理分野における主要課題に対する最先端の技術を概説する．

キーワード：多重音解析，調・和音推定，リズム解析，ビート解析，調波打楽器音分離，楽曲構造解析

1. ま　え　が　き

波形データである音楽音響信号から音楽的に意味のある情報を取り出す音楽信号処理技術が実現できれば音楽を対象とした様々な情報処理が可能になってくる．音楽信号処理の究極の目標は，計算機に音楽を人間と同じように聴き，理解し，演奏し，編曲し，創作する能力を備えさせることである．

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．音楽信号処理の研究の重要性が認識され始めた頃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，音楽には音声にない様々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急速に発展してきている．

音楽と音声との相違点には以下のようなものが挙げられる．まず，音声においては音韻（音声における音色）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．例えば音声で音韻系列がそっくり変われば異なる言語メッセージになるように，音楽で旋律，リズム，和声がそっくり変われば異なる曲になる．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．通常，音声信号処理（音声認識など）では対象となる音声は一つであり，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成る．後述する多重音解析は，多重音から各楽音の基本周波数（音高に相当する物理量）を推定するための技術である．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．もちろん音声にも広い意味でリズムがあり，コミュニケーションにおいて非言語的役割を担っているが，前述のとおり音楽においてリズムは旋律と和声と並んで重要な言語的役割を担っている．リズム・ビート解析は文字どおり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，音楽は大域的な繰返し構造や共通構造を有している．例えば，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．

次章以降で，音楽信号処理の重要トピックを紹介し，

亀岡弘和　正員　日本電信電話株式会社NTTコミュニケーション科学基礎研究所
E-mail　kameoka.hirokazu@lab.ntt.co.jp
中村友彦　東京大学大学院情報理工学系研究科システム情報学専攻
E-mail　Tomohiko_Nakamura@ipc.i.u-tokyo.ac.jp
高宗典玄　東京大学大学院情報理工学系研究科システム情報学専攻
E-mail　norihiro_takamune@ipc.i.u-tokyo.ac.jp
Hirokazu KAMEOKA, Member (NTT Communication Science Laboratories, NIPPON TELEGRAPH AND TELEPHONE CORPORATION, Atsugi-shi, 243-0198 Japan), Tomohiko NAKAMURA, and Norihiro TAKAMUNE, Nonmembers (Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, 113-8656 Japan).
電子情報通信学会誌　Vol.98 No.6 pp.467-474 2015年6月

最新の研究の動向を解説する．

2. 多重音解析・音源分離

ヴァイオリンなどのようにピッチ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言う．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付ける最も重要な情報の一つでこれを自動獲得できれば自動採譜，楽音分離，音楽検索など様々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．

多重音が対象となる場合，各楽音に分離さえできれば単一音の基本周波数推定問題に帰着するため，多重音解析の問題は音源分離の問題とも密接に関係している．このことを明快にするため，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について考えよう．もし信号が純音の場合，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を基本周波数とみなすのは頑健な方法ではない（図1(d)）．以上から，基本周波数を推定するためには，スペクトルピークのような限られた情報だけでなく，対象とする音の信号波形やスペクトル構造の全体を手掛かりにした方法が必要になる．しかし，複数の信号が混合されて観測される音響信号には，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掛かりが得られないのである．したがって，音源分離の問題が解かれない限り個々の基本周波数を推定するのは容易ではない，ということになる．一方，個々の音の基本周波数が既知であれば音源分離の問題は解きやすくなるわけなので，音源分離の問題を解く手掛かりになり得る基本周波数の情報が，音源分離の問題が解かれない限り高精度に求められない，といういわゆる「鶏と卵」の状況に陥るのである．

このことから，多重音解析の問題は音源分離とセットで考えられることが多い．例えば従来，信号中で最も優勢な基本周波数を推定するステップと対応する基本周波数成分と調波成分を対象の信号から減算するステップを反復する逐次推定アプローチが提案されている[(1)]．これに対し，全音源の基本周波数を一挙に推定しようという同時推定アプローチも提案されている．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（例えば文献(2)～(5)）．筆書らも，音源分離と基本周波数推定の問題をパラメトリックな調波構造モデル及び時間周波数構造モデルを用いた同時最適化問題として定式化し，音源分離に相当するステップと基本周波数パラメータを推定するステップを反復するアルゴリズムを提案している[(6)～(8)]．このほかにも多重音から基本周波数を推定する手法は膨大にあるので，より詳しい動向についてはほかの著書文献(9)，(10)を参照されたい．

これまで紹介した手法では各音源のスペクトル構造に関する先験知識を利用することが前提となっていたが，このような知識が事前に得られない場合もある．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返し生起する

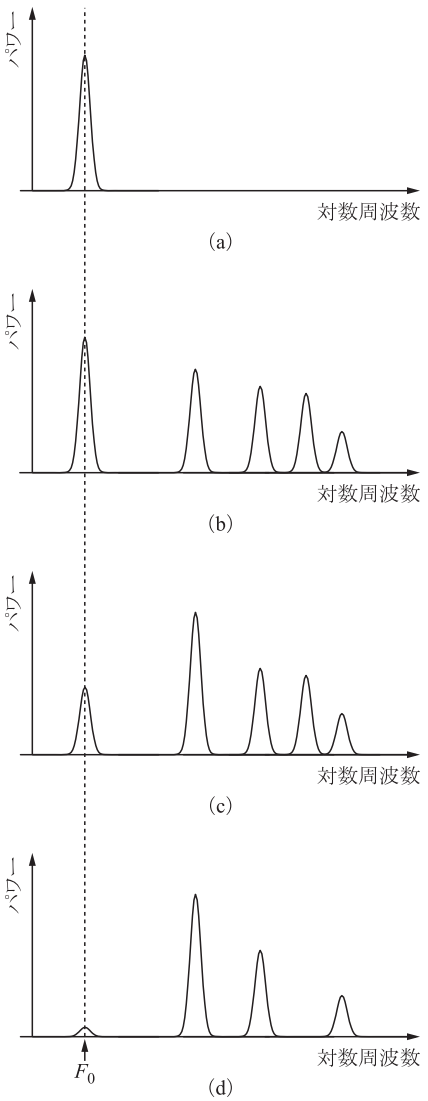

図1　基本周波数推定の問題

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[11]．例えば図2の(a)のようなスペクトルの音が(b)のような音量軌跡で鳴っていたとしよう．その音のスペクトログラムは，(a)を縦ベクトル，(b)を横ベクトルとして両者の積により得られる行列で表せる．また，(c)のようなスペクトルの音が(d)のような音量軌跡で鳴っていた場合も同様である．スペクトログラムが加法的であると仮定すると，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．逆に言えば，観測された多重音のスペクトログラムを行列 $\boldsymbol{Y}$ とし，$\boldsymbol{Y}$ を二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及び音量軌跡の情報が得られることを意味する．ここで，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を置かなければならない点に注意が必要である．このことから，このアプローチは非負値行列因子分解（NMF : Non-negative Matrix Factorization）と呼ばれる．また，楽音の調波構造において基本周波数と各倍音の間隔が対数周波数軸上でシフト不変となる性質により，対数周波数スペクトルのテンプレートと音高の音量分布との畳込みにより多重音スペクトルを表現できることに着目した拡張手法[12],[13]も提案されており，MIREXと呼ぶ音楽検索に関連する要素技術の国際コンテストにおいて2013年に本手法が最高性能を示している[14]．なお，本手法は筆者らのSpecmurt法と呼ぶ多重音解析法[15]に非負値制約を置いたものとみなせる．NMFやその拡張版について詳しく知りたい読者は文献(16)を参照されたい．

以上の多重音解析・音源分離技術により，例えば各楽音の音高や音量をユーザが自分好みに操作可能な音楽再生システムを実現することができる[17]．

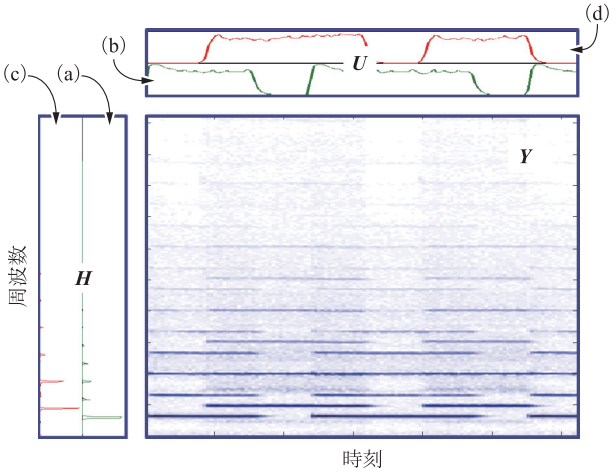

図2　NMFによるスペクトログラムの分解

3. 調・和音推定

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．Musical Instrument Digital Interface（MIDI）信号を入力とする場合もあるが，本章では音響信号入力にのみ言及する．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り，これらの情報は楽曲の類似度計算，構造解析，ジャンル認識などの有用な手掛かりとなり得る．また，人手で調・和音のラベルを付与するには労力が掛かることから[18]，調・和音推定技術の実現への期待は大きい．

人間の音高に関する知覚はtone heightとchromaの2要素に分離でき，音名（例えば，C4の音高ならばC）が同一の2音は同一のchromaに属すると知覚される[19]．そのため，音名の組合せが同一であれば異なる音高であっても同一の和音とみなすことができる．このような特徴量として，スペクトログラムを音名ごとにオクターブ間で足し合わせたクロマグラム（Pitch Class Profile, PCPとも呼ばれる）[20],[21]がしばしば用いられる．図3を見ると，同一の和音の部分が非常に似ていることが確認できる．

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掛か

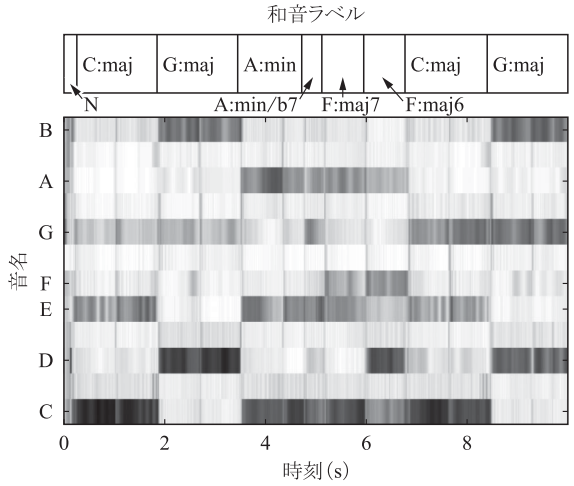

図3　クロマグラムと対応する和音ラベルの例　　楽曲としてLet It Be（The Beatles）を用いた．色の濃い部分が値が高く，淡い部分が値が低い．

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，文献(22)の研究を筆頭にHMM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多く提案されている．また，和音間の遷移のしやすさは調に強く依存するため，調の情報を取り入れる方法が提案されている[23]．更に，繰返し構造や拍位置などの楽曲構造に関する情報も導入し，ダイナミックベイズネットワークを用いて転調を扱いつつ和音推定を行う手法も提案されている[24),(25]．筆者らの研究室でも音声認識とのアナロジーに着目し，HMMに基づく調推定[26]や和音推定[27]の手法を開発してきた．ほかにも，機能和声理論に基づいたモデル，和音系列の構造（カデンツなど）を用いた手法も提案している[28]．調・和音推定に関する研究は膨大にあるのでより詳しい動向については文献(29)を参照されたい．

4. リズム解析

多重音解析により推定された各楽音の音高や発音時刻（オンセット），消音時刻（オフセット）やMIDI信号からテンポや各楽音の音価（2分音符や8分音符といった楽譜上の音の長さ）を推定する問題をリズム解析と呼ぶ．各楽音の音高とリズムの情報が得られれば楽譜のクエリを用いて楽曲の分類や類似度計算ができるようになるため，リズム解析技術は自動採譜だけでなく楽曲検索・推薦に応用することができる．

観測上の時間における音の長さは，楽譜上の音の長さとテンポの積によって決まるため，所与の演奏に対し楽譜とテンポの組合せは無数に存在する．更に人間の演奏にはテンポやオンセット，オフセットに揺らぎがあるため，リズム解析は，様々な不確定性の下で入力演奏を最も良く説明するもっともらしいリズムとテンポの組合せを見つける問題となる．

以上のようにリズム解析の問題は不良設定の逆問題であるにもかかわらず，人間はテンポ変動やオンセット・オフセットの揺らぎを含んだ演奏を聴いても，楽譜やテンポを認識することができる．恐らくこれは，人間は楽譜やテンポに対して先験知識を持ち，その先験知識に沿うようにもっともらしい楽譜とテンポを推定できているためであろう．筆者らの研究室ではこの考え方をヒントに，楽譜とテンポに関する先験知識を確率的生成モデルの形で立て，各楽音のオンセット情報から楽譜とテンポを確率的逆問題として同時推定する手法を提案している[30)~(32)]．また最近は，多旋律の各声部やパートといった同期構造をモデル化するために確率文脈自由文法を二次元拡張したモデル（図4）を用いた手法[31),(32]を検討している．

5. 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二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することができるようになれば，それぞれの音量や音色を自由に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言う．

図5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傾向にある一方で，打楽器音は周波数成分が周波数方向に平行に連なる傾向にある．前者は，同一音高

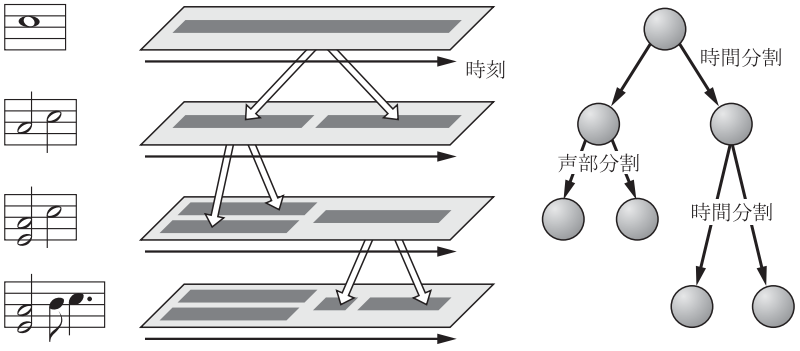

図4　二次元確率文脈自由文法による楽譜の生成モデル

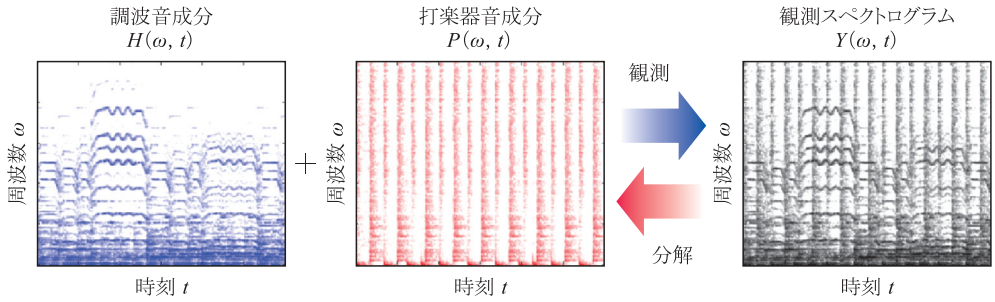

図 5　調波打楽器音分離の問題

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，広帯域に及ぶスペクトルが打叩時に急しゅん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[33]．この手法のアイデアと基本原理は筆者が大学院時代に考案したものであるが，その後当時の研究室スタッフや後輩たちの手によってアルゴリズムもプログラムも高度に洗練化され，現在フリーソフトウェアとして文献(34)で公開されている．このアプローチは国内外で関心を集めており，多くの研究者たちにより改良・拡張手法が現在進行形で検討されている（例えば文献(35)，(36)）．

なお，HPSS 以外のアプローチとしては，打楽器音のスペクトルテンプレートを用いた手法[37]，独立部分空間解析（ISA : Independent Subspace Analysis）と呼ぶ信号分解法を用いた手法[38]，前述の NMF に基づくアプローチ[39]などが提案されている．

6. 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 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，テンポ解析と言う．前述のリズム解析が解かれればこれらの問題も解決するため，ビート・テンポ解析はリズム解析の下位概念に相当する．ただし，リズム解析ではその問題の難しさゆえに現状はピアノの独奏など比較的単純な演奏が対象となることが多いのに対し，ビート・テンポ解析ではオーケストラやポピュラー音楽など音響的により複雑な演奏が対象となることが多い．拍の情報は楽曲のリズム構造の認識や類似度を計算するためのアライメントやリズムによる楽曲のジャンル分類などに有用であることから，音楽検索技術において重要な役割を果たす．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，打楽器音や各ノートが拍位置で発音されることが多いこと，などが本問題の解決の手掛かりとなる．

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，拍は音楽に内在するリズムであるため，全ての拍位置でノートが発音されるわけではないし，拍位置以外でノートが発音されることもあるため，たとえ各ノートのオンセット（発音開始）時刻が既知の MIDI 信号を対象とした場合でもビートやテンポを推定するのは容易ではない．MIDI 信号の場合は各ノートのオンセットは既知であるが，音響信号を対象とした場合，まず各時刻において発音された音があったらしいかを表すオンセット特徴量を抽出する必要がある．詳細は文献(40)に譲るが，これまでオンセット特徴量として，Spectral Flux[41]や Phase Deviation[42]，近年では深層学習により得られる特徴量[43]などが用いられている．以上の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ため，短時間フーリエ変換（STFT : Short-Time Fourier Transform）を用いた手法[44]やエージェントベースの手法[45]，マルコフモデルを用いた手法[46]，動的計画法を用いた手法[47]などが提案されている．

7. 楽譜追跡

楽譜追跡とは，所与の楽譜を参照しつつ，演奏音響信号から実時間で現在の楽譜上の位置（楽譜位置）を推定する技術である．演奏音響信号だけでなく MIDI 信号を入力とすることも多い．この技術の目的は，人間の演奏に自動で伴奏を同期させ再生する自動伴奏や自動譜めくり[48],[49]などを実現することであり，文献(50)，(51)の研究以来これまで活発に研究されてきた[52]．特に自動伴奏では，演奏が入力されない区間でも伴奏を演奏に合わせ再生するため，テンポ推定まで含めて楽譜追跡と呼ぶこともある．

人間の演奏には，テンポや強弱，誤り，弾き直し・弾き飛ばしなど様々な不確定要素が存在し，同一の楽譜に基づく演奏であっても毎回異なる．このような楽譜から一意に予測することが難しい不確定要素を含む演奏に対していかに頑健に追従するかが，楽譜追跡の重要な課題である．楽譜追跡の重要な手掛かりである音高情報を演奏音響信号から捉えるために様々な音響特徴量が提案されている．単純に，STFT によって得られたスペクトルを用いる場合もあるが，半音単位で中心周波数が並べられた定 Q 変換によるスペクトルやそれを加工した特徴量もよく使用される[53]．多くの従来研究では，これらの不確定要素を確率的に生成されたとみなし，演奏を確率モデルにより記述するアプローチを採用している．このアプローチの利点は，不確定要素も含む演奏の統計的な性質を確率に反映でき，演奏サンプルがあればパラメータ学習できる点である．確率モデルとして，HMM やその拡張[54),(55]パーティクルフィルタ[56),(57]，条件付確率場[58]などが用いられている．

筆者らの研究室では，特に練習やリハーサルなどに多く含まれる誤りや弾き直し・弾き飛ばしが任意に起き得る状況でも追跡可能な楽譜追跡アルゴリズムを提案している[59]（デモ動画像が文献(60)で視聴可能）．

8. 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と呼ぶ音楽的な構造単位に分割し，それぞれのセグメントを音楽的に同一の機能を持つカテゴリーに分類する問題である．ここで，セグメントは分割された音楽音響信号を指し，カテゴリーへの分類が行われていないものを指す．カテゴリーの例として，ポピュラー音楽のさびや A メロなど（セクション）やソナタ形式の楽曲の提示部や展開部などがある．各セグメントにセクション名を割り当て得る，さびの自動検出[61]や楽曲のサムネイル（試聴用音源など）自動生成[62]など様々なアプリケーションに役立つ．Songle[63]と呼ばれる音楽鑑賞 Web サービスでは，セクション単位で再生をスキップする機能の実現に構造解析技術が用いられている．

構造を基礎付ける音楽の構成要素の関係性は，様々な基準によって作られる．文献(64)では，基礎的な基準として「新規性」，「同質性」，「繰返し構造」が挙げられている．例えば，フィルインなど突然の変化が生じれば，異なるセクションが始まったことが分かる（新規性）．調やテンポ，楽器編成などがほとんど変化せず一貫している区間は，同一のセクションが演奏されていることが分かる（同質性）．ポピュラー音楽の 1 番と 2 番のさびなどのように，旋律や和音系列，リズムパターンなどが繰返し用いられていれば同一のセクションとみなせる（繰返し構造）．様々な音楽の構成要素（旋律，リズム，和音，楽器編成など）について，統合的に判断して構造を推定する必要があり，構造解析を自動化する上での課題となっている．

楽器編成に対応する音色の特徴量としてメル周波数ケプストラム係数（MFCC : Mel-Frequency Cepstral Coefficient），旋律や和音を表す音高にはクロマベクトルがよく用いられる．テンポやリズムパターンを表現するため，オンセット検出の結果などをベースとした特徴量が提案されている[64]．楽曲中で類似した部分を得るために，ある時刻と全時刻との特徴量の類似度を計算し，それを並べた行ベクトルを縦に連結して得られる自己類似度行列が多くの従来研究で用いられている（図 6）[64),(65]．単純なフレーム同士の類似度だけでなく，周りのフレームの情報[66]や高次の時間構造[67]を用いた類似度の計算も提案されている．

自己類似度行列上で，近辺の類似度が高いブロック状の箇所（図 6 の赤線で囲まれた部分）は同質性が高く，対角上にあるブロック同士の継ぎ目（青線同士の交点）で新規性が高い．非対角成分上で対角に走る線が繰返し構造を表している．そのため，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[68]．ここでは，あくまでセグメントだけが得られることに注意されたい．ほかにも，同質性に着目し得られたセグメントをクラスタリングする方法[69),(70]や，繰返し構造を表す非対角成分上の対角に走る線を動的計画法[71]や画像処理の手法[61]を用いて得る方法も提案されている（詳細は文献(64)参照）．また，筆者らの研究室では，得られた繰返し構造や大域的な冗長性を利用した，新たな音楽音響信号の符号化方式を提案している[72]．

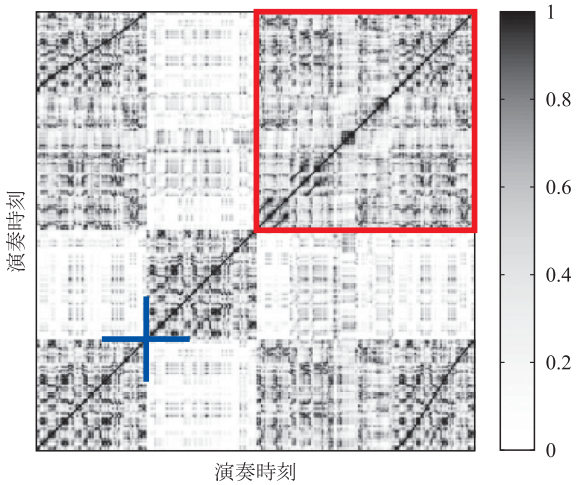

図 6　SM Toolbox[65]を用いて計算した類似度行列の例

9. むすび

本稿では，多重音解析，調・和音認識，リズム解析，調波打楽器音分離，ビート検出，楽譜追跡，楽曲構造解析など，音楽音響信号処理における重要課題に対する最先端の技術を紹介した．音楽音響信号処理に関する動向，信号処理論や機械学習理論の基礎，ツール類についてより詳しく調べたい読者は文献(10)，(16)，(29)，(40)，(73)，(74)などを参照されたい．

謝辞　本稿では，東京大学の客員連携講座・亀岡研究室及び同大学嵯峨山研究室における過去の発表資料から図の素材を一部使用した．関係各位に感謝する．

文　献

(1) A. Klapuri, "Multiple fundamental frequency estimation based on harmonicity and spectral smoothness," IEEE Trans. ASLP, vol. 11, no. 6, pp. 804-816, 2003.

(2) M.G. Christensen and A. Jakobsson, "Multi-pitch estimation," in Synthesis Lectures on Speech Audio Process., Morgan and Claypool, San Rafael, CA, 2009.

(3) R. Badeau, V. Emiya, and B. David, "Expectation-maximization algorithm for multi-pitch estimation and separation of overlapping harmonic spectra," Proc. ICASSP'09, pp. 3073-3076, 2009.

(4) C. Yeh, "Multiple fundamental frequency estimation of polyphonic recordings," Ph.D. Thesis, Univ. Pierre et Marie Curie, Paris 6, 2008.

(5) M. Goto, "A real-time music-scene-description system : Predominant-F0 estimation for detecting melody and bass lines in real-world audio signals," Speech Commun., vol. 43, no. 4, pp. 311-329, 2004.

(6) H. Kameoka, T. Nishimoto, and S. Sagayama, "Separation of harmonic structures based on tied Gaussian mixture model and information criterion for concurrent sounds," Proc. ICASSP'04, pp. 297-300, 2004.

(7) H. Kameoka, "Statistical approach to multipitch analysis," Ph.D. Thesis, The University of Tokyo, 2007.

(8) H. Kameoka, T. Nishimoto, and S. Sagayama, "A multipitch analyzer based on harmonic temporal structured clustering," IEEE Trans. ASLP, vol. 15, no. 3, pp. 982-994, 2007.

(9) A. de Cheveigné, "Multiple F0 estimation," in Computational Auditory Scene Analysis : Principles, Algorithms and Applications, D.-L. Wang and G.J. Brown eds., IEEE Press/Wiley, 2006.

(10) Signal Processing Methods for Music Transcription, A. Klapuri and M. Davy, eds., Springer, New York, 2006.

(11) P. Smaragdis and J.C. Brown, "Non-negative matrix factorization for music transcription," Proc. WASPAA'03, pp. 177-180, 2003.

(12) P. Smaragdis, B. Raj, and M. Shashanka, "Sparse and shift-invariant feature extraction from non-negative data," Proc. ICASSP'08, pp. 2069-2072, 2008.

(13) E. Benetos, S. Cherla, and T. Weyde, "An efficient shift-invariant model for polyphonic music transcription," Proc. MML'13, pp. 1-4, 2013.

(14) http://www.music-ir.org/mirex/wiki/2013 : Main_Page

(15) S. Saito, H. Kameoka, T. Nishimoto, and S. Sagayama, "Specmurt analysis of multi-pitch music signals with adaptive estimation of common harmonic structure," Proc. ISMIR'05, pp. 84-91, 2005.

(16) 亀岡弘和，"非負値行列因子分解とその音響信号処理への応用，" 日本統計学会誌，vol. 44, no. 2, pp. 383-407, 2015.

(17) 亀岡弘和，ルルー・ジョナトン，大石康智，柏野邦夫，"Music factorizer：音楽音響信号をノート単位で編集できるインタフェース，" 情処学音楽情報科学研報，2009-MUS-81, no. 9, pp. 1-6, 2009.

(18) J.A. Burgoyne, J. Wild, and I. Fujinaga, "An expert ground truth set for audio chord recognition and music analysis," Proc. ISMIR'11, pp. 633-638, 2011.

(19) R.N. Shepard, "Circularity in judgments of relative pitch," J. Acoust. Soc. Am., vol. 36, no. 12, pp. 2346-2353, 1964.

(20) G.H. Wakefield, "Mathematical representation of joint time-chroma distributions," Proc. SPIE'99, ASPAAI, vol. 3807, pp. 637-645, 1999.

(21) T. Fujishima, "Real-time chord recognition of musical sound : A system using common lisp music," Proc. ICMC'99, pp. 464-467, 1999.

(22) A. Sheh and D.P. Ellis, "Chord segmentation and recognition using EM-trained hidden Markov models," Proc. ISMIR'03, pp. 185-191, 2003.

(23) K. Lee and M. Slaney, "Acoustic chord transcription and key extraction from audio using key-dependent HMMs trained on synthesized audio," IEEE Trans. ASLP, vol. 16, no. 2, pp. 291-301, 2008.

(24) M. Mauch, "Automatic chord transcription from audio using computational models of musical context," Ph.D. Thesis, Queen Mary University of London, 2010.

(25) M. Mauch and S. Dixon, "Simultaneous estimation of chords and musical context from audio," IEEE Trans. ASLP, vol. 18, no. 6, pp. 1280-1289, 2010.

(26) 齊藤翔一郎，西本卓也，嵯峨山茂樹，"Specmurt 分析と HMM を用いた音楽音響信号の調認識，" 音響論集，pp. 757-756, 2005.

(27) Y. Ueda, Y. Uchiyama, T. Nishimoto, N. Ono, and S. Sagayama, "HMM-based approach for automatic chord detection using refined acoustic features," Proc. ICASSP'10, pp. 5518-5521, 2010.

(28) 上田　雄，小野順貴，嵯峨山茂樹，"機能和声モデルによる音楽信号からの和声推定，" 情処学音楽情報科学研報，2010-MUS-86, no. 13, pp. 1-6, 2010.

(29) M. McVicar, R. S.-Rodriguez, Y. Ni, and T.D. Bie, "Automatic chord estimation from audio : A review of the state of the art," IEEE Trans. ASLP, vol. 22, no. 2, pp. 556-575, 2014.

(30) 武田晴登，西本卓也，嵯峨山茂樹，"確率モデルによる多声音楽演奏の MIDI 信号のリズム認識，" 情処学論，vol. 45, no. 3, pp. 670-679, 2004.

(31) H. Kameoka, K. Ochiai, M. Nakano, M. Tsuchiya, and S. Sagayama, "Context-free 2D tree structure model of musical notes for bayesian modeling of polyphonic spectrograms," Proc. ISMIR'02, pp. 307-312, 2012.

(32) 高宗典玄，亀岡弘和，嵯峨山茂樹，"2 次元 LR パーサによる音楽演奏 MIDI 信号からの自動採譜，" 音響論集，pp. 1039-1042, 2014.

(33) 宮本賢一，ルルー・ジョナトン，亀岡弘和，小野順貴，嵯峨山茂樹，"スペクトログラムの滑らかさの異方性に基づく調波音・打楽器音の分離，" 音響論集，pp. 903-904, 2008.

(34) http://hil.t.u-tokyo.ac.jp/software/HPSS/

(35) D. FitzGerald, "Harmonic/percussive separation using median filtering," Proc. DAFx'10, 2010.

(36) J. Driedger, M. Müller, and S. Disch, "Extending harmonic-percussive separation of audio signals," Proc. ISMIR'14, pp. 611-616, 2014.

(37) K. Yoshii, M. Goto, and H.G. Okuno, "Automatic drum sound description for real-world music using template adaptation and matching methods," Proc. ISMIR'04, pp. 184-191, 2004.

(38) C. Uhle, C. Tiddmar, and T. Sporer, "Extraction of drum tracks from polyphonic music using independent subspace analysis," Proc. ICA'03, pp. 843-847, 2003.

(39) M. Helen and T. Virtanen, "Separation of drums from polyphonic music using non-negative matrix factorization and support vector machine," Proc. EUSIPCO'05, 2005.

(40) J.P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M.B. Sandler, "A tutorial on onset detection in music signals," IEEE Trans. Speech Audio Process., vol. 13, no. 5, pp. 1035-1047, 2005.

(41) P. Masri, "Computer modelling of sound for transformation and synthesis of musical signal," Ph.D. Thesis, University of Bristol, 1996.

(42) J.P. Bello, C. Duxbury, M. Davies, and M. Sandler, "On the use of phase and energy for musical onset detection in the complex domain," IEEE Signal Process. Left., vol. 11, no. 6, pp. 553-556, 2004.

(43) S. Böck and M. Schedl, "Enhanced beat tracking with context-aware neural networks," Proc. DAFx'11, pp. 1-5, 2011.

(44) P. Grosche, M. Muller, and F. Kurth, "Cyclic tempogram—A mid-level tempo representation for musicsignals," Proc. ICASSP'10, pp. 5522-5525, 2010.

(45) M. Goto, "An audio-based real-time beat tracking system for music

with or without drum-sounds," J. New Music Res., vol. 30, no. 2, pp. 159-171, 2001.
(46) N. Whiteley, A.T. Cemgil, and S.J. Godsill, "Bayesian modelling of temporal structure in musical audio," Proc. ISMIR, pp. 29-34, 2006.
(47) D.P. Ellis, "Beat tracking by dynamic programming," J. New Music Res., vol. 36, no. 1, pp. 51-60, 2007.
(48) 電子楽譜フェアリー，iPhone, iPad 向けアプリ，http://www.fairy-score.com/
(49) Tonara, Sheet Music Viewer That Listens To You, http://tonara.com/
(50) R.B. Dannenberg, "An on-line algorithm for real-time accompaniment," Proc. ICMC'84, pp. 193-198, 1984.
(51) B. Vercoe, "The synthetic performer in the context of live performance," ICMC'84, pp. 199-200, 1984.
(52) N. Orio, S. Lemouton, D. Schwarz, and N. Schnell, "Score following : State of the art and new developments," Proc. NIME'03, pp. 36-41, 2003.
(53) C. Joder, S. Essid, and G. Richard, "Learning optimal features for polyphonic audio-to-score alignment," IEEE Trans. ASLP, vol. 21, no. 10, pp. 2118-2128, 2013.
(54) P. Cano, A. Loscos, and J. Bonada, "Score-performance matching using HMMs," Proc. ICMC'99, pp. 441-444, 1999.
(55) C. Raphael, "Automatic segmentation of acoustic musical signals using hidden Markov models," IEEE Trans. Pattern Anal. Mach. Intell., vol. 21, no. 4, pp. 360-370, 1999.
(56) Z. Duan and B. Pardo, "A state space model for online polyphonic audio-score alignment," Proc. ICASSP'11, pp. 197-200, 2011.
(57) T. Otsuka, K. Nakadai, T. Takahashi, T. Ogata, and H.G. Okuno, "Real-time audio-to-score alignment using particle filter for coplayer music robots," EURASIP J. Appl. Signal Process., vol. 2011, pp. 1-13, 2011.
(58) C. Joder, S. Essid, and G. Richard, "A conditional random field framework for robust and scalable audio-to-score matching," IEEE Trans. ASLP, vol. 19, no. 8, pp. 2385-2397, 2011.
(59) S. Sagayama, T. Nakamura, E. Nakamura, Y. Saito, H. Kameoka, and N. Ono, "Automatic music accompaniment allowing errors and arbitrary repeats and jumps," POMA, vol. 21, 035003, pp. 1-11, 2014.
(60) https://www.youtube.com/watch?v=KgnR2BzrafU, https://www.youtube.com/watch?v=fW6VKiC4k34
(61) M. Goto, "A chorus section detection method for musical audio signals and its application to a music listening station," IEEE Trans. ASLP, vol. 14, no. 5, pp. 1783-1794, 2006.
(62) M.A. Bartsch and G.H. Wakefield, "Audio thumbnailing of popular music using chroma-based representations," IEEE Trans. MM, vol. 7, no. 1, pp. 96-104, 2005.
(63) Songle, http://songle.jp/
(64) J. Paulus, M. Müller, and A. Klapuri, "State of the art report : Audio-based music structure analysis," Proc. ISMIR'10, pp. 625-636, 2010.
(65) M. Müller, N. Jiang, and H.G. Grohganz, "SM Toolbox : MATLAB implementations for computing and enhancing similarity matrices," Proc. AES'14, London, UK, 2014.
(66) M. Müller and F. Kurth, "Enhancing similarity matrices for music audio analysis," Proc. ICASSP'06, pp. 231-236, 2006.
(67) T. Jehan, "Creating music by listening," Ph.D. Thesis, Massachusetts Institute of Technology, 2005.
(68) J. Foote, "Automatic audio segmentation using a measure of audio novelty," Proc. ICME'00, pp. 452-455, 2000.
(69) M. Cooper and J. Foote, "Summarizing popular music via structural similarity analysis," Proc. WASPAA'03, pp. 127-130, 2003.
(70) J.J. Aucouturier, F. Pachet, and M. Sandler, ""The way it Sounds" : timbre models for analysis and retrieval of music signals," IEEE Trans. Multimed., vol. 7, no. 6, pp. 1028-1035, 2005.
(71) M. M Goodwin and J. Laroche, "A dynamic programming approach to audio segmentation and speech/music discrimination," Proc. ICASSP'04, pp. 309-312, 2004.
(72) 藏　悠子，鎌本　優，小野順貴，嵯峨山茂樹，"動的計画法に基づく音楽構造解析とその音楽信号符号化への応用，" 音響論集，pp. 1049-1050, 2012.
(73) 後藤真孝，緒方　淳，"音楽・音声の音響信号の認識・理解研究の動向，" コンピュータソフトウェア，vol. 26, no. 1, pp. 4-24, 2009.
(74) M. Müller, D.P.W. Ellis, A. Klapuri, and G. Richard, "Signal processing for music analysis," IEEE J.-STSP, vol. 5, no. 6, pp. 1088-1110, 2011.

（平成 27 年 2 月 6 日受付　平成 27 年 3 月 9 日最終受付）

亀岡　弘和（正員）

平 14 東大・工・計数卒．平 19 同大学院博士課程了．同年日本電信電話株式会社入社．NTT コミュニケーション科学基礎研究所配属．平 23 東大大学院情報理工学系研究科客員准教授．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，情報処理学会，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．

中村　友彦

平 23 東大・工・計数卒．平 25 同大学院情報理工学系研究科修士課程了．同年から同大学院博士課程在籍．音楽情報処理・音響信号加工に関する研究に従事．SICE Annual Conference 2011 International Award 受賞．

高宗　典玄

平 24 東大・工・計数卒．同年から同大学院博士前期課程に在籍．平 27-04 同大学院博士課程．音楽情報処理，音響信号処理に関する研究に従事．