

小特集—音楽音響における最近の話題—

音楽情報処理の最近の研究*

後藤真孝(産業技術総合研究所)**・平田圭二(NTT コミュニケーション科学基礎研究所)**

43.75.-z

1. 音楽情報処理が実世界と結び付く

最近、音楽情報処理の産業応用が急激に拡大している。古くから研究されていた、シンセサイザに代表される楽音合成、MIDI シーケンサを基盤としたデスクトップミュージック(DTM)、各種作曲・演奏・録音支援のような音楽家向けの技術は実用化され、既に音楽制作に不可欠なものとなっている。そうした従来技術に対し、近年、音楽家でないエンドユーザが直接恩恵を受けるような技術に注目が集まっている。例えば、コンパクトディスク(CD)から吸出した(リップングした)音響信号を圧縮し、個人の計算機上で多数の楽曲を容易に処理できるようになった。Apple社のiPodのように、一万曲を携帯型音楽プレーヤに入れて、いつでもどこでも手軽に持ち歩けるようになった。

こうした動向が生まれた背景として、まず、計算機のハードウェアの進歩(処理速度の向上、メモリ・ハードディスクの小型大容量化)や、インターネットの普及、安価な音響入出力デバイスが標準装備されるようになったことが挙げられる。加えて、1992年に規格化されたMP3(MPEG Audio Layer-3)が1990年代後半から普及し、エンドユーザからの需要に応じるビジネスが成立するようになったことも挙げられる。このような流れは2000年代前半になって更に加速しており、MP3以外にもOgg Vorbis, MPEG-4 AAC, WMA(Windows Media Audio)等の圧縮方式が提案され、インターネット経由で音楽配信をする企業も続々と出現している。

音楽家でないエンドユーザは、一般に音符や和

声といった音楽知識にあまり詳しくなく、創作に対するこだわりも少ない。むしろ、聴きたい音楽や楽曲のある部分を的確に探し出して聴くようなことに、興味を持っている。最近の音楽情報処理もそのようなエンドユーザの需要を反映し、処理対象が個々の楽曲の内部(音符や和音等)から楽曲全体やその集合へと大きくなり、音楽知識がなくとも利用可能な音楽システムも活発に研究されている。このような動向を支える技術には、例えば、音楽の類似度を計算し的確に検索したり分類したりする技術、他の人達がどのような音楽を聴いているかを参考に音楽を選ぶ技術、音楽を扱うための優れたインタフェース技術がある。

本稿では、こうした新しい研究の流れを中心に、従来とは違った切口で音楽情報処理の最近の研究を紹介していく。

2. 楽曲の集合を扱う

従来の研究は個々の楽曲の内部を対象としていたのに対し、ここ10年で、音楽情報検索(MIR: Music Information Retrieval)という、楽曲を単位として大量の楽曲の集合に対する検索、分類、管理に取り組む研究分野が立ち上がってきた。国際会議ISMIR(International Conference on Music Information Retrieval)が2000年以降毎年開催される等、活発に研究されている。様々な課題が取り組まれているが、ここでは単なる書誌情報(CDDbによる曲名やアーティスト名等)に基づくテキスト検索でなく、音響信号に基づく検索に焦点を当てて、以下の三つに分類して紹介する。

2.1 メロディを扱う研究: ハミング検索(QBH: Query by Humming)

聞いたことのある曲を「ラララー」等のように口ずさむと、その曲名を検索できる方法である。メロディの歌唱やハミングを検索キーとして、そのメロディを持つ楽曲を検索する。検索キーについ

* Recent studies on music information processing.

** Masataka Goto (National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, 305-8568)

*** Keiji Hirata (NTT Communication Science Laboratories, Kyoto, 619-0237)

ては、音が外れた歌唱のような誤りへの対処、調やテンポの違いの吸収等が課題となる。具体的な手法は、データベースとして、メロディのみ [1-5]、楽曲全体の標準 MIDI ファイル (SMF) [6-8]、楽曲全体の音響信号 [9, 10] のどれを用いるかによっても異なる。メロディのみの場合には、検索キーとの類似度を直接求めればよいが、SMF の場合には、どのトラックがメロディかを推定してから、類似度を求めなければならない。更に音響信号の場合には、混合音中のメロディとの類似度を求める必要があり、より実現が難しい。

2.2 楽曲の断片を扱う研究：断片を含む楽曲の検索

町中で流れている音楽の曲名を知りたいときに、携帯電話でその一部分を録音すると、曲名を検索できる方法である。楽曲の断片を検索キーとして、その断片を含む楽曲を検索する。高速化と共に、雑音や伝送路の歪等の音響的な変動をいかに吸収するかが課題となる。ベクトル量子化されたパワースペクトルの形状のヒストグラムに基づく時系列アクティブ探索法 [11] や、パワースペクトルのピークの出現パターンに基づく方法 [12] などが提案されている。

2.3 楽曲全体を扱う研究：楽曲間の類似度に基づく検索

ある曲が気に入っているときに、それに似た曲調の曲を探ることができる方法である。楽曲を検索キーとして、それに類似した楽曲を検索する。そのためには、楽曲間の類似度を、楽曲中の音色 (パワースペクトル形状) [13, 14]、リズム [14-17]、変調スペクトル [18]、歌声 [19] 等の様々な特徴に基づいて定義する必要がある。類似度は検索以外にも重要であり、類似度に基づく楽曲の自動分類 (音楽ジャンル、曲調の分類) [14, 17-20] 等も研究されている。楽曲同士が似ていることを適切に類似度に反映することは難しく、次の 3 章で紹介する音楽音響信号理解と共に、今後のさらなる研究が期待される。

3. 音楽音響信号を理解する

音楽の音響信号を理解する研究も過去 10 年間に大きく展開した。それ以前は、音響信号の個々の構成音を分離・抽出する音源分離や、その情報から楽譜を作成する自動採譜が主流であった。しかし、

人間が音楽を理解することについて再考し、「人間は、構成音の分離信号や楽譜に基づいて音楽を理解しているわけではない」という立場から音楽理解を目指す考え方 (3.2 節、音楽情景記述 [21-23]) が 1997 年に提案され [24]、ビートトラッキング、メロディ抽出、音楽構造分析等の研究テーマが次々と生まれた。

こうした音楽音響信号理解が大きく発展した背景には、音響信号を扱う手法とハードウェアの両者の進歩が挙げられる。10 年前は高速フーリエ変換 (FFT) を実時間で計算することさえ困難だったが、今日、その計算時間は無視できる程速くなった。こうした処理能力の向上により、従来は思い付かなかったような力業の計算が可能になり、それが新手法の発想にも大きく影響を与えている。特に、各種統計的手法は広く導入され、隠れマルコフモデル (HMM) 等の確率モデルに基づく手法や、最尤推定、ベイズ推定等を活用した様々な手法が提案されている。

3.1 音源分離・音高推定

自動採譜は、1970 年代から研究されている歴史の長いテーマだが、単旋律から、単一楽器の多重音、複数種類の楽器による混合音へと、対象の複雑さを徐々に増しながら進歩してきた。それに伴い、音源分離、音高 (基本周波数) 推定といったより特化した課題への取組みにシフトしてきた。

研究事例が多く、網羅的な紹介はできないため、ここでは過去 10 年に登場した新しい考え方を中心に紹介する。1994 年に、柏野らによって確率モデルに基づく手法が導入され、処理モデル OPTIMA [25, 26] が実装された。周波数成分、単音、和音の階層構造をグラフィカルモデルで関連付け、それらの階層的な関係を考慮しながら最も尤らしい解釈を求められる点に新規性があった。1999 年には、後藤によって音源数を仮定しない音高推定手法 PreFEst [21, 23, 27] が提案された。この手法では、高調波構造の形状を表す確率分布をあらゆる音高に対して用意し、それらの混合分布 (重み付き和) として入力周波数成分をモデル化する。そのモデルパラメータを、EM (Expectation-Maximization) アルゴリズムを用いて最大事後確率推定することで、混合音の各構成音の音量 (重み) と高調波構造の形状が同時に求められる点に新規性があった。原理的に非高調波構造にも拡張

でき [23, 27], 混合音理解のフレームワークと位置付けられる。

ほかに、最も優勢な音高の推定とその高調波成分の除去を繰り返すことで、混合音の構成要素を順次求めていく手法 [28] や、音の波形の重み付き和として時間領域でモデル化し、MCMC (Markov chain Monte Carlo) アルゴリズムによって同時発音数、各音を構成する周波数成分数、音高、音量等のモデルパラメータを推定する手法 [29]、楽譜が (局所的な) あるテンポで演奏されて波形が生成される過程を、グラフィカルモデルで音符とテンポ、波形を関連付けてモデル化し、そのすべてを推定する手法 [30]、周波数成分を高調波構造の拘束下でクラスタリングする問題と定式化し、赤池情報量規準 (AIC) を最小にするクラスタ数 (音源数) を求めながら、各クラスタの重心 (音高) と重み (音量) を推定する手法 [31] 等が提案されている。

3.2 音楽情景記述

従来の研究の多くが、音楽音響信号から個々の音符や個々の構成音の分離信号を得ることを目指していたのに対し、音楽情景記述 (music scene description) [21–23] では、音楽的に訓練されていない「しろうと」の音楽理解の実現を目指し、メロディ、ベース、ビート、サビ、フレーズの繰り返し、楽曲構造、楽器の音色等の音楽演奏中の情景を分析・理解した結果を記述する点が特徴的である。以下では、個々の記述を得る手法を紹介する。メロディ、ベース 複数の楽器音が混在した混合音中のメロディ、ベースの音高推定は、上記で紹介した PreFEst [21, 23, 27] を帯域制限して適用することで、1999 年に初めて実現された。その後、ベースの音高推定に関しては、Hainsworth らによって別の手法も提案された [32]。

ビート 階層的なビート構造 (小節) を得るビートトラッキング、小節推定は、MIDI 信号に対しては 1980 年代から研究されていたものの、混合音に対しては後藤らの 1994 年からの一連の研究 [24, 33–36] で初めて実現された。そして、この影響を受けて、テンポ変化に対応できるビートトラッキング手法 [37] が Scheirer によって提案された。その後も、より制約の少ない多様な手法 [38–40] が提案されている。MIDI 信号を対象とした手法 [41, 42] も近年大きく進展した。

サビ、フレーズの繰り返し、楽曲構造 2000 年代に入り、楽曲中で繰り返される類似区間 (フレーズ等) の検出に基づく考え方が登場し、楽曲の代表的な部分 (多くの場合サビ) を一箇所切り出す手法 [43–45] や、主要な部分を残して短くする音楽要約手法 [46, 47]、すべてのサビ区間を網羅的に検出する手法 RefraiD [48] 等が提案された。この中で、RefraiD はサビ検出に焦点を当てており、転調の有無に関わらず各サビ区間の開始点、終了点を求めることができる特長を持っている。

楽器の音色 1990 年代後半から、混合音中の個々の音の楽器名を、音高と同時に推定する音源同定手法 [49, 50] や、ドラム音の発音時刻を推定する手法 [33, 34, 36, 51, 52] が実現されるようになった。単独演奏音を対象とした音源同定手法 [53–58] も様々な観点から研究されている。

4. 楽曲のメタ情報を扱う

聴きたい音楽を的確に探し出して聴くというエンドユーザの需要に直接的に応えるため、楽曲の作曲者や演奏者の情報、楽曲に対するリスナの嗜好等のメタ情報を、楽曲聴取や検索のために抽出して活用する研究が盛んになっている。

4.1 利 用

他の人達がどういう音楽を聴いているかを参考に、音楽を選ぶことができるエンドユーザにとって便利である。インターネット上の音楽通販サイト (amazon 等) や音楽批評サイト (allmusic 等) では、ユーザの評判や感想、購買履歴等のメタ情報が日々蓄積されている。それらの情報に協調フィルタリング (collaborative filtering) 技術を適用することで、ユーザに対するアーティストやアルバムの推薦 [59–61] や、プレイリスト (元来は放送やコンサート等での演奏曲目リストの意味であったが、ここではメディアプレーヤ等の再生曲目リストを指す) の提案 [62–65] 等、購買意欲を促進するサービスが実現できる。

4.2 抽 出

協調フィルタリングだけでは未知の新曲に対する対処が困難なため、内容に基づくフィルタリング (content-based filtering) と併用することが望ましい [61]。そのためには、3 章で述べたような音楽理解技術が重要であり、その結果を用いるこ

とで、楽曲の音響的特徴や内容に基づくより適切な推薦やプレイリスト作成が可能になる。

こうしたメタ情報と音響的特徴を併用した例として、Whitman らは、あるアーティストの楽曲の音響的特徴と、そのアーティスト名を含む WWW ページ中の語句の統計情報を組合せ、アーティストのスタイルの判定 [20] やアーティストの識別 [66] を実現した。Ellis らは、音楽批評サイト上の類似アーティストのリストや、エンドユーザの持っている楽曲コレクション、アーティスト名を含む WWW ページ中の語句の統計情報からメタ情報を抽出し、アーティスト間の類似度がどの程度求められるかを検討した [67]。更に Berenzweig らは、こうしたメタ情報に基づく類似度と、音響信号に基づく類似度とを比較した [68]。

4.3 記述と標準化

楽譜情報のみの符号化方法については既に十分実用レベルに達しているが [69]、メタ情報を含む記述法や標準化については、現在幾つかの XML ベースの提案が並立している状況にある。例えば、楽譜情報を含む記号レベルの音楽記述の提案には、MusicXML [70] や WEDELMUSIC [71] がある。また、メロディの形状やパワースペクトルの統計情報等の音楽音響信号の内容に関するメタ情報の記述に関しては、MPEG-7 Audio [72] が標準化されている。今後、これらに準拠した研究開発が一層活発になると予想される。

5. エンドユーザが使うことを考える

音楽知識に詳しくないエンドユーザが音楽を思い通りに扱えるようにするには、既存の音楽家向けのツールでは不十分であり、新たなインタフェースについての研究が重要となる。

5.1 実世界指向

敷居が低く自然に使えるインタフェースを実現するために、実世界に存在する物体をそのままインタフェースとして利用する方法が考えられる。ここで、その物の従来からの操作方法も踏襲していることが大切である。そのような実世界指向の例として、musicBottles [73] では、楽曲を演奏する各楽器パートを異なるガラス瓶に対応付け、各ガラス瓶の栓を開けている間だけそのパートの音が鳴るような音楽再生インタフェースを実現した。FieldMouse [74] は、バーコードリーダーのような

ID タグ認識装置と、マウスのような相対位置検出装置を組み合わせた入力デバイスであり、選曲やボリューム変更を表す ID タグを読み取って空中で動かすことで、それらの操作を可能にした。

既存の音楽試聴機やメディアプレーヤでは、エンドユーザが試聴のためにサビの部分だけを聴きたい場合、早送りボタンを何度も押して自分でサビを探す手間がかかる。SmartMusicKIOSK [75] では、前述の RefraiD [48] を応用して新たに「サビ出し」ボタンを追加し、自動検出されたサビの先頭へボタンひと押しで飛べる機能を実現した。これにより、CD プレーヤの「次の曲」ボタンで興味のない楽曲を容易に飛ばせるように、楽曲内部で興味のない箇所を容易に飛ばせるようになった。

5.2 擬音語

今後エンドユーザが様々な音楽システムを操作することを考えた場合、楽曲そのものの情報(メロディ、リズム等)を全く入力しないで済ませることは難しい。そのときには、ハミング(2.1 節)や以下に述べる擬音語が有力な入力手段となる。「ストロン音楽」表記システム [76] では、「ドレミーソツミレド」のようにテキストで旋律を記述すると、それを計算機に演奏させることができる。口(くち)ドラム検索手法 [77] では、「ドンタンドタン」のように自然に口ずさんだドラムパートの音声を認識して、楽曲検索に用いることを目指している。

5.3 コミュニケーションツール

エンドユーザは通常、あまり創作に対して興味を持たない傾向にあるが、例えば orkut のようなソーシャルネットワーキングツールの中に、手軽な音楽創作のための支援機能を埋め込むことができれば、音楽もエンドユーザによるコミュニケーション手段の一つとして利用できるようになる可能性がある。

ユーザ間のコミュニケーションを重視したシステムの例として、CosTune [78] では、上着やズボンに貼り付けられた異なる音を出す複数のパッドにリズムカルに触れて演奏する楽器によって、近傍の人々と無線ネットワーク接続して合奏することを可能にした。Music Resonator [79] では、アノテーションの付加された楽曲断片をユーザが加工・編集し、それをユーザ間で共有することで、手軽に共同音楽制作をできるようにした。Remote-

GIG [80] では、遅延の大きいインターネットを介して、遠隔地にいるユーザ同士が12小節のブルース等をリアルタイムに合奏することを可能にした。お互いにコード進行の1周期分(数十秒)だけ遅れた相手の演奏を聴くことで、ネットワーク遅延を吸収した合奏を実現した。

6. 音楽情報処理は今後更に発展する

音楽情報処理の研究分野では、本稿で紹介した以外にも様々な魅力的なテーマが活発に研究されている。ここでは音響信号処理に関連したトピックを中心にあげたが、楽譜やMIDIを含めた記号処理の研究も進展しており、例えば、記号レベルのメロディの類似度を計算する研究 [81] や、演奏に付けられた表情と音楽的な構造の対応についての研究 [82] がある。こうした記号処理と音響信号処理の融合はまだ不十分で、今後の重要な課題の一つである。そうした融合によって補い合うことで、適切な記号接地 (symbol grounding) に基づく記号処理や、抽象的な意味計算 (semantic computing) に踏み込んだ音響信号処理が可能になり、多面的な音楽の意味を反映した計算が実現できると考えられる。

音楽情報処理を研究するための環境も整備されつつある。2000~2001年に研究目的で共通に利用できる世界初の著作権処理済み音楽データベース「RWC 研究用音楽データベース」 [83] が構築され、各種手法の比較・評価や、データベースに基づく学習、研究成果の对外発表で、音楽が利用しやすくなった。他の研究分野では、古くから多様な共通データベースが構築されて研究の進展に大きく貢献してきたが、音楽情報処理の研究分野でも同様の進展が期待できる。

10年前には、音楽情報処理は「遊び」でなく「研究」であることを認めてもらうための説明が必要だったのに対し、今日ではそれが重要な研究分野であることが常識となった。大型プロジェクトが次々と生まれ、関連国際会議の種類も年々増え、研究者も増え続けているこの分野の、今後のさらなる発展を期待したい。

文 献

[1] T. Kageyama, K. Mochizuki and Y. Takashima: Melody retrieval with humming. *Proc. ICMC 1993*, pp. 349–351 (1993).

[2] A. Ghias, J. Logan, D. Chamberlin and B.C. Smith: Query by humming: Musical information retrieval in an audio database. *Proc. ACM Multimedia 95*, pp. 231–236 (1995).

[3] T. Sonoda, M. Goto and Y. Muraoka: A WWW-based melody retrieval system. *Proc. ICMC 1998*, pp. 349–352 (1998).

[4] S. Pauws: CubyHum: A fully operational query by humming system. *Proc. ISMIR 2002*, pp. 187–196 (2002).

[5] T. Sonoda, T. Ikenaga, K. Shimizu and Y. Muraoka: The design method of a melody retrieval system on parallelized computers. *Proc. WEDEL-MUSIC 2002*, pp. 66–73 (2002).

[6] J. Shifrin, B. Pardo, C. Meek and W. Birmingham: HMM-based musical query retrieval. *Proc. JCDL 2002*, pp. 295–300 (2002).

[7] N. Hu and R.B. Dannenberg: A comparison of melodic database retrieval techniques using sung queries. *Proc. JCDL 2002*, pp. 301–307 (2002).

[8] R.B. Dannenberg, W.P. Birmingham, G. Tzanetakis, C. Meek, N. Hu and B. Pardo: The MUSART testbed for query-by-humming evaluation. *Proc. ISMIR 2003*, pp. 41–47 (2003).

[9] T. Nishimura, H. Hashiguchi, J. Takita, J.X. Zhang, M. Goto and R. Oka: Music signal spotting retrieval by a humming query using start frame feature dependent continuous dynamic programming. *Proc. ISMIR 2001*, pp. 211–218 (2001).

[10] J. Song, S.Y. Bae and K. Yoon: Mid-level music melody representation of polyphonic audio for query-by-humming system. *Proc. ISMIR 2002*, pp. 133–139 (2002).

[11] K. Kashino, T. Kurozumi and H. Murase: A quick search method for audio and video signals based on histogram pruning. *IEEE Trans. Multimedia*, 5, 348–357 (2003).

[12] A. Wang: An industrial-strength audio search algorithm. *Proc. ISMIR 2003*, pp. 7–13 (2003).

[13] J.-J. Aucouturier and F. Pachet: Music similarity measures: What's the use? *Proc. ISMIR 2002*, pp. 157–163 (2002).

[14] G. Tzanetakis and P. Cook: Musical genre classification of audio signals. *IEEE Trans. Speech Audio Process.*, 10, 293–302 (2002).

[15] J. Paulus and A. Klapuri: Measuring the similarity of rhythmic patterns. *Proc. ISMIR 2002*, pp. 150–156 (2002).

[16] J. Foote, M. Cooper and U. Nam: Audio retrieval by rhythmic similarity. *Proc. ISMIR 2002*, pp. 265–266 (2002).

[17] S. Dixon, E. Pampalk and G. Widmer: Classification of dance music by periodicity patterns. *Proc. ISMIR 2003*, pp. 159–165 (2003).

[18] M.F. McKinney and J. Breebaart: Features for audio and music classification. *Proc. ISMIR 2003*, pp. 151–158 (2003).

[19] W.-H. Tsai, H.-M. Wang, D. Rodgers, S.-S. Cheng and H.-M. Yu: Blind clustering of popular music recordings based on singer voice characteristics. *Proc. ISMIR 2003*, pp. 167–173 (2003).

[20] B. Whitman and P. Smaragdis: Combining musical and cultural features for intelligent style detection. *Proc. ISMIR 2002*, pp. 47–52 (2002).

[21] M. Goto: A real-time music scene description system: Detecting melody and bass lines in audio signals. *Working Notes of the IJCAI-99 Workshop on*

- Computational Auditory Scene Analysis*, pp.31–40 (1999).
- [22] M. Goto: Music scene description project: Toward audio-based real-time music understanding. *Proc. ISMIR 2003*, pp.231–232 (2003).
- [23] 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.* (2004).
- [24] M. Goto and Y. Muraoka: Real-time rhythm tracking for drumless audio signals — chord change detection for musical decisions —. *Working Notes of the IJCAI-97 Workshop on Computational Auditory Scene Analysis*, pp.135–144 (1997).
- [25] K. Kashino: *Computational Auditory Scene Analysis for Music Signals*. PhD thesis, University of Tokyo (1994).
- [26] K. Kashino, K. Nakadai, T. Kinoshita and H. Tanaka: Organization of hierarchical perceptual sounds: Music scene analysis with autonomous processing modules and a quantitative information integration mechanism. *Proc. IJCAI-95*, pp.158–164 (1995).
- [27] M. Goto: A predominant-F0 estimation method for polyphonic musical audio signals. *Proc. ICA 2004*, pp. II-1085–1088 (2004).
- [28] A.P. Klapuri: Multiple fundamental frequency estimation based on harmonicity and spectral smoothness. *IEEE Trans. Speech Audio Process.*, 11, 804–816 (2003).
- [29] M. Davy and S.J. Godsill: Bayesian harmonic models for musical signal analysis. *Bayesian Stat.*, 7, 105–124 (2003).
- [30] A.T. Cemgil, B. Kappen and D. Barber: Generative model based polyphonic music transcription. *Proc. WASPAA 2003*, pp.181–184 (2003).
- [31] H. Kameoka, T. Nishimoto and S. Sagayama: Extraction of multiple fundamental frequencies from polyphonic music using harmonic clustering. *Proc. ICA 2004*, pp. I-59–62 (2004).
- [32] S.W. Hainsworth and M.D. Macleod: Automatic bass line transcription from polyphonic music. *Proc. ICMC 2001*, pp.431–434 (2001).
- [33] M. Goto and Y. Muraoka: A beat tracking system for acoustic signals of music. *Proc. ACM Multimedia '94*, pp.365–372 (1994).
- [34] M. Goto: *A Study of Real-time Beat Tracking for Musical Audio Signals*. PhD thesis, Waseda University (1998).
- [35] M. Goto and Y. Muraoka: Real-time beat tracking for drumless audio signals: Chord change detection for musical decisions. *Speech Commun.*, 27, 311–335 (1999).
- [36] M. Goto: An audio-based real-time beat tracking system for music with or without drum-sounds. *J. New Music Res.*, 30, 159–171 (2001).
- [37] E.D. Scheirer: Tempo and beat analysis of acoustic musical signals. *J. Acoust. Soc. Am.*, 103, 588–601 (1998).
- [38] S. Dixon: Automatic extraction of tempo and beat from expressive performances. *J. New Music Res.*, 30, 39–58 (2001).
- [39] S. Hainsworth and M. Macleod: Beat tracking with particle filtering algorithms. *Proc. WASPAA 2003*, pp.91–94 (2003).
- [40] A.P. Klapuri, A.J. Eronen and J.T. Astola: Analysis of the meter of acoustic musical signals. *IEEE Trans. Speech Audio Process.* (2004).
- [41] A.T. Cemgil and B. Kappen: Monte carlo methods for tempo tracking and rhythm quantization. *J. Artif. Intell. Res.*, 18, 45–81 (2003).
- [42] H. Takeda, T. Nishimoto and S. Sagayama: Automatic rhythm transcription of multiphonic MIDI signals. *Proc. ISMIR 2003*, pp.263–264 (2003).
- [43] B. Logan and S. Chu: Music summarization using key phrases. *Proc. ICASSP 2000*, pp. II-749–752 (2000).
- [44] M.A. Bartsch and G.H. Wakefield: To catch a chorus: Using chroma-based representations for audio thumbnailing. *Proc. WASPAA '01*, pp.15–18 (2001).
- [45] M. Cooper and J. Foote: Automatic music summarization via similarity analysis. *Proc. ISMIR 2002*, pp.81–85 (2002).
- [46] G. Peeters, A.L. Burthe and X. Rodet: Toward automatic music audio summary generation from signal analysis. *Proc. ISMIR 2002*, pp.94–100 (2002).
- [47] R.B. Dannenberg and N. Hu: Pattern discovery techniques for music audio. *Proc. ISMIR 2002*, pp.63–70 (2002).
- [48] M. Goto: A chorus-section detecting method for musical audio signals. *Proc. ICASSP 2003*, pp. V-437–440 (2003).
- [49] K. Kashino and H. Murase: A sound source identification system for ensemble music based on template adaptation and music stream extraction. *Speech Commun.*, 27, 337–349 (1999).
- [50] J. Eggink and G.J. Brown: A missing feature approach to instrument recognition in polyphonic music. *Proc. ICASSP 2003*, pp. V-553–556 (2003).
- [51] A. Zils, F. Pachet, O. Delerue and F. Gouyon: Automatic extraction of drum tracks from polyphonic music signals. *Proc. WEDELMUSIC 2002*, pp.179–183 (2002).
- [52] K. Yoshii, M. Goto and H.G. Okuno: Automatic drum sound description for real-world music using template adaptation and matching methods. *Proc. ISMIR 2004* (2004).
- [53] K.D. Martin: *Sound-Source Recognition: A Theory and Computational Model*. PhD thesis, MIT (1999).
- [54] A. Eronen and A. Klapuri: Musical instrument recognition using cepstral coefficients and temporal features. *Proc. ICASSP 2000*, pp. II-753–756 (2000).
- [55] J.C. Brown, O. Houix and S. McAdams: Feature dependence in the automatic identification of musical woodwind instruments. *J. Acoust. Soc. Am.*, 109, 1064–1072 (2001).
- [56] M.A. Casey: Reduced-rank spectra and minimum-entropy priors as consistent and reliable cues for generalized sound recognition. *Proc. CRAC 2001* (2001).
- [57] P. Herrera, A. Yeterian and F. Gouyon: Automatic classification of drum sounds: A comparison of feature selection methods and classification techniques. *Proc. ICMAI 2002*, pp.69–80 (2002).
- [58] T. Kitahara, M. Goto and H.G. Okuno: Musical instrument identification based on F0-dependent multivariate normal distribution. *Proc. ICASSP 2003*, pp. V-421–424 (2003).
- [59] U. Shardanand and P. Maes: Social information filtering: Algorithms for automating “word of mouth.” *Proc. CHI '95*, pp.210–217 (1995).
- [60] W.W. Cohen and W. Fan: Web-collaborative filtering: Recommending music by crawling the Web.

- Proc. WWW9* (2000).
- [61] A. Uitdenbogerd and R. van Schyndel: A review of factors affecting music recommender success. *Proc. ISMIR 2002*, pp. 204–208 (2002).
- [62] M. Alghoniemy and A.H. Tewfik: A network flow model for playlist generation. *Proc. ICME 2001* (2001).
- [63] S. Pauws and B. Eggen: PATS: Realization and user evaluation of an automatic playlist generator. *Proc. ISMIR 2002*, pp. 222–230 (2002).
- [64] B. Logan: Content-based playlist generation: Exploratory experiments. *Proc. ISMIR 2002*, pp. 295–296 (2002).
- [65] J.-J. Aucouturier and F. Pachet: Scaling up music playlist generation. *Proc. ICME 2002* (2002).
- [66] B. Whitman: Semantic rank reduction of music audio. *Proc. WASPAA 2003*, pp. 135–138 (2003).
- [67] D.P. Ellis, B. Whitman, A. Berenzweig and S. Lawrence: The quest for ground truth in musical artist similarity. *Proc. ISMIR 2002*, pp. 170–177 (2002).
- [68] A. Berenzweig, B. Logan, D. P. Ellis and B. Whitman: A large-scale evaluation of acoustic and subjective music similarity measure. *Proc. ISMIR 2003*, pp. 99–105 (2003).
- [69] E. Selfridge-Field, Ed.: *Beyond MIDI* (The MIT Press, Cambridge, Mass., 1997).
- [70] M. Good: Representing music using XML. *Proc. ISMIR 2000* (2000).
- [71] P. Bellini and P. Nesi: WEDELMUSIC format: An XML music notation format for emerging applications. *Proc. WEDELMUSIC 2001*, pp. 79–86 (2001).
- [72] ISO/IEC JTC1/SC29/WG11 Moving Picture Experts Group: Information technology — multimedia content description interface — part 4: Audio, 15938-4:2002 (2002).
- [73] H. Ishii, A. Mazalek and J. Lee: Bottles as a minimal interface to access digital information. *Proc. CHI 2001*, pp. 187–188 (2001).
- [74] T. Masui and I. Sii: Real-world graphical user interfaces. *Proc. HUC 2000*, pp. 72–84 (2000).
- [75] M. Goto: SmartMusicKIOSK: Music listening station with chorus-search function. *Proc. UIST 2003*, pp. 31–40 (2003).
- [76] T. Masui: Music composition by onomatopoeia. *Proc. IWEC 2002*, pp. 297–304 (2002).
- [77] T. Nakano, J. Ogata, M. Goto and Y. Hiraga: A drum pattern retrieval method by voice percussion. *Proc. ISMIR 2004* (2004).
- [78] K. Nishimoto, T. Maekawa, Y. Tada, K. Mase and R. Nakatsu: Networked wearable musical instruments will bring a new musical culture. *Proc. ISWC 2001*, pp. 55–62 (2001).
- [79] K. Hirata, S. Matsuda, K. Kaji and K. Nagao: Annotated music for retrieval, reproduction, and exchange. *Proc. ICMC 2004* (2004).
- [80] M. Goto, R. Neyama and Y. Muraoka: RMCP: Remote music control protocol — design and applications —. *Proc. ICMC 1997*, pp. 446–449 (1997).
- [81] W.B. Hewlett and E. Selfridge-Field, Eds.: *Melodic Similarity: Concepts, Procedures, and Applications* (The MIT Press, Cambridge, Mass., 1998).
- [82] *Proc. ICAD 2002 Rencon Workshop* (2002).
- [83] M. Goto: Development of the RWC music database. *Proc. ICA 2004*, pp. I-553–556 (2004).



後藤 真孝

1998年早稲田大学大学院理工学研究科博士後期課程修了。博士(工学)。同年、電子技術総合研究所(2001年に産業技術総合研究所に改組)に入所し、現在に至る。2000~2003年科学技術振興事業団さきがけ研究21研究員を兼任。音楽情報処理、音声言語情報処理などに興味を持つ。情報処理学会山下記念研究賞(音楽情報科学・音声言語情報処理)、電気関係学会関西支部連合大会奨励賞、WISS2000論文賞・発表賞、日本音響学会粟屋潔学術奨励賞・ポスター賞、日本音楽知覚認知学会研究選奨、インタラクシオン2003ベストペーパー賞各受賞。日本音響学会、情報処理学会、電子情報通信学会、日本音楽知覚認知学会、ISCA各会員。



平田 圭二

1987年東京大学大学院工学系研究科情報工学専門課程博士課程修了。工学博士。同年、NTT基礎研究所入社。1990~1993年財新世代コンピュータ技術開発機構(ICOT)に出向、その間、並列推論マシンの研究開発に従事。1999年よりNTTコミュニケーション科学基礎研究所、現在に至る。音楽知プログラミング、インタラクシオンに興味を持つ。情報処理学会平成13年度論文賞受賞。情報処理学会、日本人工知能学会、日本ソフトウェア科学会各会員。