

■2群 (画像・音・言語) - 9編 (音楽情報処理)

1章 基礎・境界

(執筆者：片寄晴弘) [2011年7月 受領]

■概要■

本章では、音楽情報処理という領域を俯瞰していくにあたり、音楽に関連した用語、複数の技術領域に関連した基礎技術・境界領域について紹介する。

音楽に関連した用語といえば、音楽学という専門領域があり、そこで、様々な音楽用語の定義がなされている。それらの網羅的な紹介は専門書(例えば、*The Grove Dictionary of Music and Musicians*, Oxford University Press など)に譲り、本章前半では、音楽情報処理に関連して重要なものに絞って、音楽心理、音楽認知など、自然科学に近い立場から紹介を試みる。

2000年代以降インターネットの社会的な普及が進み、音楽の流通の仕方もそれまでのモノを媒体としたものからデータとして配信する形へと大きく変遷し、ネット上に大量のデータ音楽が存在・流通していることを前提とした新たな音楽情報処理関連技術が進展した。章後半では、これらに関連して重要性を増したキーワード、複合領域について紹介する。章最後で、技術領域全般に関連する音楽生成言語・環境について紹介する。

■2群-9編-1章

1-1 主要音楽用語（音楽心理学の立場から）

（執筆著：津崎 実）[2011年6月 受領]

1-1-1 音楽心理学の方向性

音楽心理学という学問領域があるのかと改めて問われるとその答えに窮するところがある。その一つには音楽心理学という領域に期待されている学術的な貢献が二面性をもっているからであると考えられる。元々ギリシャ哲学の時代から音楽は学問の中で重要な位置を占めており、音楽に対する思弁を行うことには学問的な意義があった。音楽というものが人間の内部に働きかけるもの、つまり今日的な表現をすれば、音楽という刺激を受けることにより何らかの情報処理がなされると考えられる以上、当然のことながら現代の心理学の中でも音楽に関する情報処理を扱うことは外れな研究テーマではない。その一方で、そのような学術的探求によって何を明らかにしたいのかについての焦点は研究者により異なってくるであろう。非常に乱暴な二分法をすれば、音楽に対する探求を最終目的として心理学的な手法や知識を応用しようとする立場と、その反対に、心理学的な問題（知覚、認知、知覚運動協応、記憶、創造）を理解するための一つの手がかりとして音楽を材料として扱う立場があると思う。前者の立場はむしろ心理音楽学という方が適切なのかも知れないが、実際のところ優れた研究は音楽に対する理解も心理学に対する理解もともに進展させるものであるべきである。しかしながら、研究を進めるうえで、特に先行研究の成果を理解するうえで上述したような二つの立場があることを意識しておくことは重要である。音楽学や音楽理論を研究の背景とした研究者の研究成果の中には心理学のほかの分野で既に確立されていた知見に従った解釈学的なものが多くなるであろう。つまり心理学的には新規の知見に乏しい。反対に、実証科学としての心理学の訓練を受けた研究者の成果の中には、音楽学では従来理論的にだけ議論されてきた内容を、実証することに重点を置いた研究が多くなってしまいがちである。この場合は音楽理論的な新規性は乏しくなる。これらの研究にはもちろんそれぞれの価値が存在する。ただし、理論的に言われてきただけのことを実証したデータを詳細に吟味すれば、理論の一部とは不整合を来すような実験事実も存在していることの方が多い。実験をすることの意義は、このような不整合を突破口にして新たな理論の展開への道筋を開くことにあるはずである。このような議論は音楽心理学に限らず、一般的に理論研究と実験研究の間の健全な関係に関するものであり、ここであえて繰り返すことが適切なかどうかは疑問もある。しかし、音楽に関する研究をする場合、どうしても研究者自身が好む音楽や、研究者自身の音楽観、思い入れが強く研究結果の解釈に出る傾向があると筆者は常々感じている。ここではあくまでも実証研究としての音楽心理学の立場から、主要音楽用語について解説することを目的としているので、実験事実がたとえ自分の信じるどころと相容れない結果を呈したとしてもそれを受け容れる姿勢をもつことの重要性を改めて強調しておきたい。

1-1-2 音楽の多面性

音楽心理学については、研究の焦点が音楽にあるか心理学にあるかという違いだけでなく、音楽という対象が非常に広範囲の人間の活動を包含していることによる多様性も存在している。音楽は聴く対象であり、それと同時に演奏する対象でもある。聴く場合も、初めて聴く

場合もあれば以前に聴いた曲を再び聴いて再認したり、また、後で暗唱するための記憶の対象にもなり得る。創造する対象でもあり得る。このように考えると音楽には人間のあらゆる営みの要素が入っていることになる。その中からあえて音楽というものを特別に取り出す意義がどこにあるのか、という疑問に的確な答えが出せるようになったときに音楽心理学の研究も実りの多いものとなるであろう。

1-1-3 音楽の知覚的側面

音楽の聴取にはいわゆる旋律の知覚、リズムの知覚、ハーモニーの知覚の側面があり、それぞれの観点からの研究がなされてきている。知覚研究は音楽的な刺激を入力し、それに対する反応を見るというかたちで、今日の実験心理学の基盤をなした行動主義による研究方法との親和性も高く、研究の数も多い。

(1) 旋律と音階

旋律については、まずその旋律を特徴づける特徴が何であるかを探る研究がなされている。旋律を構成する最小の単位である音（楽音）は、明確なピッチをもった音であり、旋律とはそのピッチが異なる音の系列である。その場合、もっとも原始的な特徴はピッチの上昇・下降のパターンであることが実験的にも確認されている。ただし、音楽で使われるピッチは音階として知られる離散的なピッチの集合の中から選択されている。音楽理論としてはピタゴラス学派に始まる調律論の観点から、いわば理論的にどの音の組合せを用いることが適切であるかが決められてきた。しかし、これらは必ずしも人間の聴覚にとって適切であるかという実験的な立場から決められたものではない。実験心理学としては西洋調性音楽で多用されるいわゆる全音階（1オクターブを全音、全音、半音、全音、全音、全音、半音に区切る音階）が知覚的な優位性をもっているか否かを探究する研究が行われている。また、このような体系においてはオクターブ（基本周波数の比が1:2になる関係）の等価性が前提とされている。人間の聴覚系では音響信号の振動周期に同期した神経発火が観察されるという事実はオクターブ関係にある音どうしに共通性を見出す合理的な根拠を提供している。ただし、音楽的な刺激を使った実験研究でオクターブの等価性を実証した研究例は実は数が少ない。

音階の構成については、聴覚心理学の分野では音響信号に備わる周期性を基盤とした説明がつけられると考えている研究者が多いのに対し、音楽心理学の分野では連想記憶やニューラル・ネットを基盤とした研究が主流を占めているようである。その背景には、音楽というものがいわゆる文化的な所産であり、文化による多様性を世界の様々な音楽が有しているということがある。人間の聴覚系に共通に見られる基盤に基づいた場合、文化による多様性は説明が困難となるからである。これに対してニューラル・ネットのもつ柔軟性は文化固有の変種を吸収可能である。ただし、伝承されている文化によってどのような連想記憶が構成されるかは予想可能であっても、全くの白紙状態からどのようにして現在ある形の音階が成立したかについてはニューラル・ネットに代表される学習モデルは説明の規範をもっていない。実際のところは、聴覚系にとって特徴的な手がかりとして原始的な音楽形態は成立し、その後環境要因や文化要因などの様々音響以外の拘束条件が加わることにより現在流通している音階システムが構成されたと考えらるべきであろう。この領域になると、それは音楽心理学というよりもより大きな音楽文化を取り扱うものとなってくるので、これ以上の議論はこの項目では割愛することとする。

(2) 絶対音感・音程

旋律とはピッチの違う音の連鎖であると先に述べたが、旋律についてのゲシュタルト心理学以来の疑問として、何故移調した旋律は同じ旋律だとして知覚可能であるのかということがある。これは旋律を構成する要素が単独のピッチではなく、二つのピッチの間の差、つまり音程であるという考え方に到達する。移調旋律の等価性が比較的簡単にかつ特別な訓練を特に必要もなく知覚可能であるということから、音程の知覚、別の言い方をすれば絶対音感の所有は、一般的に容易であると想定されているようである。これに対して絶対音感の能力については古くから稀少な能力とされてきており研究者の注目を集めてきた。実際には日本では西洋に比べて絶対音感保有者が多いことが知られており、これは日本における何らかの音楽教育法との関連性が指摘されている。ただし、絶対判断をすること自体は別に特別な能力でないことはほかの感覚を考えてみても想像できる。例えば、我々は虹のスペクトルを見ないことには与えられた色が赤か緑かの判断ができないということはない。更には音楽を聴くうえで真に求められるのは絶対音感ではなく相対音感であることは否定しようがなく、絶対音感研究というものが音楽心理学のどのような側面を掘り下げることになるかについては今後とも議論が必要であろう。むしろ、絶対音感研究に比べて相対音感研究の数が少ないことの方が問題である。これは先にも述べたように、誰でも旋律の異同を容易に判断できる一種の先入観から相対音感の保有は容易いものであると無批判に前提としているからであろう。日常的に接する旋律ではキー（調整の中心）が決まっており、知覚者はその中心音からの各構成音の相対的な関係を把握すると考えられる。しかし、真の意味での相対音感では中心音が任意の位置に、予想不能に移動しても音程を正確に判断する能力であるべきである。このような場面は音楽的な文脈でも滅多に存在することがなく、音楽家であってもこの判断には苦労を伴う。旋律の知覚を問う場合には、真の相対音感とはどのようなものであるかについての実証研究が更に必要であると思われる。

(3) リズム

音楽のもう一つ重要な側面にリズムがある。音楽の成立と舞踊については強い関連性があり、音楽のもつリズムには何らかの身体的な制約があると考えられてきている。これは Fraisee によって *indifference interval* と規定された時間で、およそ 600 ms がそれに相当する。人間はこの長さについては過小評価も、過大評価もしにくいとされる。このような時間の存在は何らかの意味での内的クロックの存在を示唆するものであり、内的クロックの存在については種々の観点からの研究がなされている。しかしながら、単一の内的クロックの存在を示すようなデータは得られていないのが現状である。むしろ、階層的な時定数をもったロー・パス・フィルタ群の構成により、入力された音楽音響信号に存在する拍節構造を推定する計算モデルが提案されている。このモデルは基本的にボトム・アップに動作するものであるが、出力結果は調性音楽に対する生成文法理論 (Generative Theory of Tonal Music) に基づいた規則ベースのシステムが解析した結果と非常に類似した結果を示している。

(4) ハーモニー

ハーモニーとは音楽の和声論の基礎になる概念であり、同時に複数の楽音が呈示されている場合に、そのどのような組合せが協和的に聞こえ、どの組合せが不協和的に聞こえるかに関する音楽理論的な考え方がこの概念の根底にある。この概念と密接に関連している聴覚心理学的な知見は臨界帯域に代表される聴覚の周波数分解能に関するものである。2以上の楽

音が存在するという場合は、物理的には少なくとも二つ以上の異なる周波数成分があることを意味する。最も単純な場合は、二つの異なる周波数をもつ正弦波が呈示された場合である。二つの正弦波が同じ振動体を強制振動させる場合にはその周波数差に応じた唸りが生ずる。2音間の周波数差が増加するにつれて唸りの周波数も増加し、それに伴って不協和度も増加する。しかし、周波数差が更に大きくなると聴覚システムは二つの周波数成分を完全に分解できるようになる。したがって、同じ場所が二つの周波数によって駆動されなくなり唸りは生じない。この唸りの発生する可能性と、唸りの周波数が不協和性の知覚に影響する。ただし、大半の楽器の音は調波複合音であるので、一つの音であっても基本周波数とその高調波成分から成立している。よって基本周波数差が広がっていても、その中の高調波間では唸りを生ずる組合せが発生する。音楽的な協和については基本周波数が小さな整数比関係にあるときに協和性が高いとなるのは、そのためと考えてよい。ただし、不協和と考えられる音を両耳に分けて呈示しても若干の不協和性が感じられる。この場合、二つの音は異なる耳を振動させているので唸りが生ずる可能性がないので、不協和感の起源は必ずしも聴覚末梢系に限定されるものだけではなくさそうである。

一方でハーモニーという概念、あるいは価値観そのものに対する疑問も投げかけられている。いわゆる西洋調性音楽の伝統ではハーモニーが重要視されているが、民族音楽学的な知見が集まるとハーモニーを重視するのは西洋調性音楽という特定の文化のものであり、すべての文化圏でこれが重要視されているわけではないことが分かってきた。協和性というものが人間の知覚の産物であり、知覚には人間の基本的な特性以外にも経験や学習の結果も反映されるため、ハーモニーの普遍性を議論することには困難が伴う。心理学的にはハーモニーという音楽的な価値観の妥当性を問うというよりは、その価値観をもった人間が判断を行うときの物理的な手がかりを解明し、その価値観をもつために必要だった経験や学習内容を明らかにするという視点を保つのが有益であろう。

1-1-4 音楽演奏

音楽にはそれを聴くという側面以外に演奏をするという側面が存在する。これも音が心理学のテーマとして成立し得る。それはより大きな枠では身体制御に関する心理学的研究との関連性が出てくる。音楽の熟達者が非熟達者感を感服させる能力の一つに初見演奏がある。つまり、一度も聴いたことのない楽曲の譜面をその場で手渡されて楽譜どおりの演奏を成し遂げる能力である。これにはもちろん、それまでの練習によって蓄積された様々なサブパターンの貢献があることは想像に難くないが、初見演奏の上手な演奏者は、1回ごとの凝視時間が短く、凝視点の前後の移動を頻繁に行うということが分かっている。また、与える楽譜に故意に音楽的には間違った標記を挿入しておく、楽譜どおりではなく音楽的に正しい演奏をしてしまうことも知られている。演奏者が、自分が出した音をどの程度参考にして演奏制御をしているかについても興味深い問題である。いわゆる遅延聴覚フィードバックは発話行動のときと同様におよそ0.25 s程度の遅延のときに演奏を非常に妨害することが知られている。いわゆる音響的楽器（電子楽器以外）の場合、自らが演奏する音を遮断したり改変したりすることは非常に困難を生ずるが、近年における技術的な発達により実時間の音響的変換は必ずしも不可能なものではなくなってきた。これら技術の活用によりフィードバックに関する実験を多種多様な楽器に適用することによって演奏にかかわる諸側面の解明が期待

される。

1-1-5 音楽と大脳生理学

音楽については、大脳機能においていわゆる言語中枢との局在の差があることが話題にされる。これは失語症になった患者が過去に覚えた歌を歌詞付きで流暢に歌うことができたという一事例から有名になったことであるが、実際にその後行われてきている様々な大脳生理学的な研究では、単純な機能局在性を主張することはできないという事実が積み重ねられてきている。音楽も言語も高度に複雑な作業であるために何らかの意味での機能局在は否定しがたい反面、音楽という行為がもつ複雑性から音楽の中枢がここにあるというような単純な図式を当てはめることができない。繰り返しになるが、音楽には、聴く、演奏する、楽譜を読むなどの諸相が関与しており、また演奏をする場合も、実際に動かす部位は受け持つ楽器によって様々であるわけである。その意味で、音楽という行為を研究対象とすることは非常に複雑な相手を扱うことになるわけであるが、それを成し遂げたときには人間の行為という複雑な対象に一步迫ることができるとも言えよう。

■参考文献

- 1) D. Deutsch (Ed.), "The Psychology of Music," 2nd Ed., Academic Press, 1999.
- 2) F. Lerdahl and R. Jackendoff, "A Generative Theory of Tonal Music," MIT Press, 1985.
- 3) P. Howell, I. Cross and R. West, "Musical Structure and Cognition," Academic Press, 1985.
- 4) T.J. Tighe and W.J. Dowling, "Psychology and Music: The Understanding of Melody and Rhythm," Lawrence Erlbaum Associates, 1993

■2群-9編-1章

1-2 聴覚

(執筆著：津崎 実) [2011年6月 受領]

聴覚において扱う範囲は言うまでもなく、聴覚というモダリティで処理される情報についてということになる。しかし、実際に我々が聴覚で何を知覚しているかについて改めて問い直してみると、その範囲は非常に広いことに気づく。多くの人々にとって聴覚は音声という人間のコミュニケーションを受容する器官として重要視されていることが多い。また、音楽という対象を聞くためにも聴覚の必要性はある。このような知覚対象の場合、文化や言語圏が異なると入力される音響刺激が同一であっても、知覚内容が異なるということは頻りに経験される。

これに対して、例えば聴覚器官に何らかの障害をもった場合はその克服は非常に困難である。これはヒトという生物として機能し得る聴覚の末梢器官の一部に生じた障害については、近年の様々な補綴技術の向上をもってしても補い切れない側面が残されているからである。本節では、まず音響信号を感受するためのヒトの聴覚機構の基本構造を主に解説する。

1-2-1 聴覚器官の構造

音響信号は大気中を伝播する疎密波である。外界に何らかのエネルギーの移動、例えば物体と物体の衝突などが生ずることによって、そこの生じた振動は周囲の大気へと伝わり我々の聴覚器官に伝わる。聴覚の末梢器官は、外耳、中耳、内耳と大別されている。外耳は一般に耳たぶと言われている耳介と、そこから頭部の中へ導かれる管である外耳道の部分を差し、鼓膜で次の段階の中耳と仕切られている。外耳道は大気で満たされているため、ここを伝播する振動は大気中の振動と同じく気体中を伝播する振動である。また、外耳道自体も管があるためにその管の径と長さに応じた共振特性をもっている。次の中耳は、中耳腔と呼ばれる空間に、ツチ、キヌタ、アブミと呼ばれる三つの小さな骨（耳小骨）が連結されている。ツチ骨は鼓膜に接しており、鼓膜の振動をキヌタ骨を介してアブミ骨に伝える。このアブミ骨は馬具の鐙(あぶみ)のような形状をしており、その底部が内耳に蓋をするように接している。内耳の中はリンパ液で満たされており、空気の振動はそのままでは液体である内耳の中に効率よく伝播しない。しかし、鼓膜とアブミ骨の底部の面積比の関係で鼓膜に伝わった振動は効率よく内耳内のリンパ液に伝わるという仕組みになっている。内耳は平衡感覚をつかさどる三半規管と一体になった殻状の構造をもっており、聴覚に関連するのはその内の蝸牛と呼ばれる部分である。この名称はこの部分が文字どおりカタツムリのように管が螺旋状に巻いていることから由来する。実際にはこの管は単純な一方の管ではなく、その中にもう一本の別の管が挿入されたような構造をもっており、断面を取ると3室に別れているように見える。この3室をそれぞれ前庭階、中央階、鼓室階と呼ぶ。ただし、前庭階と鼓室階は構造的には蝸牛頂(蝸牛の管の先端側)にある蝸牛孔につながっており、外リンパ液で満たされている。これに対して中央階は内リンパ液で満たされている。このため、中央階とそれ以外の二つの回の間にはイオン濃度の違いが存在しており、平衡状態における静止膜電位を生む。前庭階と中央階を仕切る膜はライスネル膜、鼓室階と中央階を仕切る膜は基底膜と呼ばれている。

1-2-2 聴覚器官の機能

ここまで解説した基底膜の挙動は受動的なものであったが、実際の基底膜は能動的な活動をするかも知られている。受容細胞として内毛細胞が存在することは先に述べたが、これと対峙するように外毛細胞という細胞が存在している。この場合、内外の区別は蝸牛の回転軸に対して内側にあるか、外側にあるかによってなされる。観察技術の進歩により最近では生きたままの検体を使って基底膜に生ずる振動の入力依存性を測定することが可能となった。その結果、従来屍体において観察されていた受動的な特性ではなく、能動的な特性、すなわち入力される信号レベルに応じてゲイン調性をするような特性が既に基底膜振動の段階で生じていることが発見されている。このような場合に、外毛細胞を選択的に破壊する薬物（カナマイシン）を投与することによって能動的な特性が消失することから、基底膜の段階で観察される能動的特性は外毛細胞の機能であることが定説化されている。

1-2-3 聴神経発火の位相固定

もう一点、蝸牛という聴覚にとってほんの入り口に過ぎない段階でその後の聴覚情報処理にとって重要な情報の符号化が行われていることに言及したい。先ほど、基底膜の振動の共振特性で入力信号の周波数分析がなされていることを説明した。ただし、入力信号に備わる周期性の符号化はこの基底膜における周波数分析だけに限定されない。基底膜の振動は内毛細胞に受容電位を生じて、それが活動電位として聴神経に伝わるわけであるが、このときの聴神経の活動の時間パターンはランダムではないことが生理学的な実験結果から明らかになっている。一般的には、発火は刺激が与えられている期間中ランダムな間隔で発生する。ところが、聴神経の発火パターンと入力された信号との関係を詳細に調べてみると、基底膜の振動の特定の位相で発火を示し、その逆相の場合には発火が抑制されるというパターンが存在することが見出された。すなわち、例えば 100 Hz の入力がある場合にはその周期である 10 ms という周期性が神経発火の活動パターンにも存在していることが分かっている。このような現象を神経活動の位相固定と呼ぶ。この位相固定が生ずる上限はほぼ 4 kHz 付近にあるということが生理学的な観察結果の示すところであるが、そのあたりが音楽で使う音符、つまり旋律を奏するための音の高低の限界と符合することなどから、位相固定した情報がピッチ情報の源泉であるというのが現在の定説である。このピッチの上限（約 4 kHz）と可聴域の上限である 20 kHz とを混同しない注意が必要である。4 kHz を超えても音としての存在はヒトの聴覚系はとらえられるわけであり、その成分の有無によって音色の違いなどは識別可能であるから、音楽を聴くうえで 4 kHz 以上が不必要であるという結論をこのことから導くべきではない。

1-2-4 聴覚初期過程の計算モデル

以上に述べてきた聴覚の初期過程については実際の音響信号を入力として稼働する機能モデルとして実装されてきている。外耳と中耳は総合的に見れば受動的なバンドパスフィルタとして見なすことができ、また内耳による周波数分析は中心周波数の異なるバンドパスフィルタ・バンクとして実装可能である。その後の神経発火に関しては生理学的にかなり忠実な模擬をするモデルも提供される一方で、機能的には半波整流と神経発火飽和を模擬した圧縮過程で模擬できる。更に、神経活動パターンに存在する周期性の検出は時間間隔ヒストグラ

ムを取ることによって模擬できる。このような表現変換技術により、従来フーリエ分析に基づくスペクトルやスペクトログラムを入力刺激の唯一の表現として考えてきた研究姿勢の誤謬に気づかせてくれることがある。典型的には基底膜における周波数分析は決してフーリエ分析のように中心周波数によらず周波数分解能が固定されたものではなく、中心周波数の上昇にほぼ比例的に周波数分析幅は広がっていく特性、つまり工学的には定 Q 型フィルタ・バンクの特性をもっている。その意味では基底膜で実装されている周波数分析はフーリエ分析的ではなく、ウェーブレット分析的であるといつていいであろう。このような聴覚機能モデルのもつ長所は、フーリエ分析的な発想法において周波数という概念の下に混乱しがちな知覚属性を比較的きれいに分離可能であるという点である。音の知覚的 3 大属性として、ラウドネス、ピッチ、音色があるということは多くの聴覚の入門書に書かれることである。ラウドネスを音響的なパワーの非線形変換としてとらえることはいいとして、ピッチは周波数によって決まるといわれ、その一方で周波数スペクトルは音色を規定するというのもいわれる。実際にはピッチを規定しているのは周波数といっても基本周波数であり、更にそれは聴覚的な実体としては信号に備わる周期性に位相固定して与えられる神経活動パターンに由来する情報なのである。これに対して、音色とは基底膜上の興奮パターンの分布によって規定されるものである。代表的な聴覚機能モデルは計算機上で実行可能な形態で提供されており、このようなシミュレータに同じ音色でピッチの違う音、あるいは違う音色でピッチの同じ音を入力した場合の出力を見比べることによって、聴覚的な表現が決してフーリエ分析に基づくような周波数スペクトルではないことが容易に理解できるはずである。

更に重要なことは聴覚系の入出力関係が、まさにその入り口の段階で既に非線形な特性をもっているということである。このような非線形性を忠実に模擬できる機能モデルの提供はまだ始まったばかりであり、その取扱いは線形モデルに比べて困難ではあるが、今後の聴覚研究を真剣に考えるうえでは自分が対象として扱う現象が線形性を仮定してよい範囲でのものであるかどうかを見極める必要があるであろう。

しかしながら、聴覚系をフィルタ・バンクとしてとらえる考え方は 19 世紀の後半から聴覚研究の中核をなしてきた考え方であり、一連のマスクングの実験はこのフィルタ・バンクとしての聴覚系の特性を心理物理学的な実験を通して測定することを目的として実施されてきた。初期はフィルタのバンド幅を超えた点に特性上の臨界が生ずるとの過程に則った臨界帯域の測定を中心に研究が進められてきた。しかし、その後、聴覚系のような自然物が構成するフィルタとして明確な臨界点が観察されるような矩形の周波数応答特性をもったものを仮定することは困難であるということから、現在ではノッチ・ノイズ、すなわち測定したい中心周波数のところにノッチをもったノイズによるマスクング実験が主流となっている。フィルタが山形の裾野を引いていればノッチの両側の雑音測定対象となるフィルタに活性化を起し、ノッチの幅を独立変数としてマスクング閾を測定することを通じて聴覚的なフィルタの形状の推定が可能となるという理屈である。

1-2-5 両耳間差

以上記述した聴覚的な処理は、単耳のみにかかわるものである。人間には二つの耳が存在する。これによって我々は音源の到来方向に関する情報を得ることが可能となっている。音源が真正面、あるいは真後ろにある場合を除いて、音源から到来する音響信号には両耳の間

で差が生ずる。一つの差は頭部がつくる音響的な「陰」による両耳間の強度差である。この場合、音響信号の周波数によって解説の度合いが異なり、周波数の高い成分ほど解説しにくいため両耳間強度差は高周波数成分においては音源到来方向の有力な手がかりとなる。これに対して、低周波数成分については両耳間の到達距離の違いに応じて生ずる両耳間時間差が手がかりとして機能し得る。周波数の低い音については神経信号の発火が位相固定して生ずることもこの能力にかかわってきている。聴覚における求心性の信号の伝達は、蝸牛神経核の後、上オリーブ複合体、側線状、下丘、内側膝状体を経て大脳皮質聴覚へと投射されているが、蝸牛神経核の出力の段階で既に反対側への投射が見られることは、両耳間の差に基づく処理を初期の段階から必要としていることを裏付ける生理学的な知見であると考えられる。

■参考文献

- 1) B.C.J. Moore, "An Introduction of the Psychology of Hearing," 5th Ed., Academic Press, 2003.
- 2) J.O. Pickles, "An Introduction of the Physiology of Hearing," 3rd Ed., Academic Press, 2008.
- 3) R.M. Warren, "Auditory Perception: A New Analysis and Synthesis," Cambridge Univ. Press, 1999.
- 4) C.J. Plack, "The Sense of Hearing," Lawrence Erlbaum Associates, 2005s.

■2 群-9 編-1 章

1-3 認知的音楽理論

(執筆者：平賀 譲) [2011年6月 受領]

「認知的音楽理論」とは人間が音楽を聴く認知過程を意識した音楽理論、とりわけ音楽構造の理論を指す。いわば音楽を「聴く側の理論」である。これに対し、和声学、対位法、楽式論などの伝統的な意味での音楽理論は、作曲など、音楽を「作る側の理論」といえる。この聴く・作るという方向性の違いは、両者の性格の違いにもつながっている。伝統的な音楽理論は既存の楽曲を規範とし、その分析を通じて理論が形成されるため、「古典和声」、「ジャズ和声」など、特定のジャンルや様式と結びついた理論となる。これに対し、認知的音楽理論では音楽知覚・認知についての直観や知見に基づく（あるいは基づこうとする）ため、普遍性のある音楽理論の性格を帯びてくる。伝統的な音楽理論ではほとんど取り上げられてこなかった、楽曲の「グループ」への分節や拍節の認識などの低レベルの認知、また旋律やリズムの理論は、認知的音楽理論にとってはむしろ中核的な課題となる。

伝統的な音楽理論は古代ギリシャや中世にまで源流をたどれるが、現在の形は18世紀以降に発達し、特に19世紀末頃から整備されてきた。これに対し、認知的音楽理論が登場してくるのは20世紀も後半のことである。その背景として、19世紀末からの音響心理・生理学、音楽心理学の発展が挙げられる。これは音楽を「人間の科学」の対象として、認知的な、聴取者の観点から眺める素地・気運となった。

音楽学の見地から認知的観点の重要性を説いたのがマイヤー (L. Meyer) で、1950年代以降、一連の著作で様々な所論が展開される^{1,2,3)}。これは時期的にはチョムスキー (N. Chomsky) の生成言語理論、認知心理学や人工知能の誕生など、認知科学が形成される時期と重なっており、時代思潮の反映と言えよう。マイヤーの理論は旋律論、リズム論、楽曲構造と類似性、音楽と情動など多岐にわたっており、認知的音楽理論の先駆的存在としてその後の研究に大きな影響を与えた。特に gap-fill melody のような旋律類型の分類や、暗意・実現 (Implication-Realization) の概念は重要である。しかし叙述のスタイルは論考・エッセイ的なものが多く、リズム論など一部を除けば、体系だった理論の構築を目指したものではない。

大規模で包括的な認知的音楽理論として代表的なものとして、ラーダール (F. Lerdahl) とジャッケンドフ (R. Jackendoff) による GTTM (Generative Theory of Tonal Music)⁴⁾ と、マイヤーの流れを汲むナムア (E. Narmour) による IRM (Implication-Realization Model)^{7,8,9)} の二つが挙げられる。これらは1980-90年代に登場したものだが、理論としての性格も目指すものも互いにかなり異なっている。このほかにも、日本でも村尾忠廣や保科洋が独自の理論を展開しており、また音楽心理学や音楽情報科学の中にも認知的音楽理論と目されるものもあるが、規模や知名度の点では上の二つが際立っている。

1-3-1 GTTM (Generative Theory of Tonal Music)

GTTM は音楽学者ラーダールと生成言語学者ジャッケンドフの共著⁴⁾のアクロニムであり、そのまま理論の通称にもなっている。名称にもあるように、言語学の生成文法理論の精神を踏まえて、楽曲の統合的な階層的構造の分析を目指した理論である。原著のほか、竹内によ

る記事がコンパクトな解説になっている⁹⁾。

GTTM は以下の四つの部分構造から構成される。

- ・ グループ構造 (Grouping Structure)
- ・ 拍節的構造 (Metrical Structure)
- ・ タイムスパン簡約 (Time-Span Reduction)
- ・ 延長的簡約 (Prolongational Reduction)

最初のグループ構造は動機や楽節のように、楽曲をまとまりのある単位に分節するもので、下位グループを統合して上位グループが得られるという意味で階層的である。グループは基本的にはボトムアップに構成され、楽曲全体が最上位グループとなる。

2 番目の拍節的構造は拍・拍子・小節に相当する時間的な単位の割当てだが、必ずしも楽譜上の小節線などに対応するわけではない。これも拍の強弱に応じてボトムアップに階層的に構成されるが、数小節程度の、比較的低レベルの範囲でしか考えない。図 1・1 にグループ構造及び拍節的構造の例を示す。楽譜下の点で表されているのが拍節的構造、その下のカッコでくくったのがグループ構造で、(A)、(B) 二つの構造があるのは、2通りのグループ構造解釈に対応している。

図 1・1 グループ構造、拍節的構造の例 (Mozart: Piano Sonata K.331)

3・4 番目の「簡約 (reduction)」とは、下位構造の代表音 (head) から上位構造の代表音 (実際には和音) を選ぶことをいう。階層を上がるにつれて、代表音が絞られていくことで、曲が「簡約」されていくわけである。GTTM ではこれを二分木で表現するため、二つの head の一方が他方に従属する、つまりトーナメント戦のように、head どうしの内「勝ち残った」方が上位に持ち上がるかたちで表される。

簡約構造の内容は、この head 間の従属関係をどう与えるかで決まる。タイムスパン簡約は隣接するグループの head の構造的な重要度を従属関係として、時間的なまとまりを組み上げていくもので、楽曲形式に対応するような静的な階層構造を表す。一方、延長的簡約の方は、主として和声進行に基づく緊張・弛緩関係を従属関係とし、曲の進行とともに展開される動的な階層構造を表す。この緊張・弛緩の関係は、グループ構造の境界を超えて、head が「延長される」かたちで影響が長スパンに及び得る。そのため解析は必然的に、トップダウン

的なものになる。図1・2にタイムスパン簡約木の例を示す。まっすぐ延長される枝が上位の簡約に持ち上げられ、行き止まりになっている枝はそれに従属している。



図1・2 タイムスパン簡約木の例 (Mozart: Piano Sonata K.331)

簡約概念そのものはGTTM 独自のものではなく、従来からの音楽分析にも存在していた。特に先行するシェンカー理論 (Schenkerian Analysis)⁵⁾では GTTM 同様、簡約階層による楽曲分析が中心となっている。しかしシェンカー理論では最上位構造 (Ursatz) が天下りに固定され、トップダウンの構成になっていること、分析方法の詳細が示されず、分析者に依存して恣意的になりがちといった問題がある。

これに対し、GTTM は四つの部分構造を導くルールシステムとして定式化されており、理論の透明性や客観性は高くなっている。ルールは構成ルール (well-formedness rule) と選好ルール (preference rule) の2種類からなる。構成ルールは構造の構成要件を、選好ルールは複数の候補から望ましい構造を選ぶ選択基準を表す。ゲームやスポーツでいえば、構成ルールはゲームそのものの規則に、選好ルールは戦術や定石に相当すると思えばよい。構成ルールはルールとしての限定力は弱いので、言語文法のように文・非文を峻別したりするのではなく、曲としてのまとまりの良さや解釈の妥当性を表現し、比較対照するための枠組といえる。

ルールは曲全体を統一的な枠組で解析できることを意図している。グループ境界の検出など、低レベルのルールはゲシュタルト原理などの認知的知見に基づいて具体性も高く、解析も音符データからボトムアップ的に行える。これに対し、高レベルのルールは繰り返し構造の検出や和声理論などに依拠し、記述も抽象的・一般的になっている。特に延長の簡約ではトップダウン的な扱いが必要であり、実際の適用は難しくなる。また二分木に限定された簡約表現が解析を不必要に複雑にしていること、そもそも階層構造によって音楽構造を律しきれるか、それが人間の音楽認知を反映しえているかといった原理的な問題もある。

1-3-2 IRM (Implication-Realization Model)

暗意・実現モデル (Implication-Realization Model: IRM) は、元々はマイヤーが発案したもので、人間が音楽を単に受動的に聴くのではなく、次に何が続くかの予測的な処理を伴って能動的に聴いているという考え方に基づいている。例えば「ドレミファ」という音列を聴いたとしよう。ここで一段落と思えば「解決 (closure)」の状態となる。そうでなければ未解決 (nonclosure) であり、何らかの後続進行が期待 (というより無意識に暗意) される。今の場合、「ソラシド」のように音階が続くというのが一番自然な暗意だろう。実際の曲ではその暗意がそのまま実現される場合もあるし、「ミレド」のように異なった進行になる場合もある。

後者の場合、進行自体はいったん解決に達する一方、暗意そのものは実現されずに持続され得る。このような「解決-未解決」、「暗意-実現」の関係によって曲 (特に旋律) の構造、更にはそれが引き起こす情動を分析しようというのが IRM の考え方である。ただし、シエンカー理論の場合と同様、実際の分析は分析者の主観に委ねられ、強引な分析も生じがちである。

これに対し、IRM の徹底した記号化・形式化を目指したのが、マイヤーの教え子でもあるナムアである。当初は旋律だけでなく、リズムや和声などにも IRM を拡大適用しようという意図が見られたが⁷⁾、90年代に発表された2部作^{8,9)}では旋律論に特化し、また元々のIRMからはかなり様変わりしている。

理論の基本的な考え方として、まず2音の音程 (音高の差) を：

- ・ 方向：上がる／下がる／等しい
- ・ 大きさ：大きい／小さい／等しい

のように質的に分類する。すると音の三つ組は、「連続して小さく上がる (P)」、「大きく上がって小さく下がる (R)」、「大きく上がって小さく上がる (IR)」、「小さく上がって元の高さに下がる (ID)」のようなパターンに分類できる (図1・3)。P, R, IR, IDなどはナムアが用いる記号で、基本的なパターンは16個、ほかに補助的な記号がある。



図1・3 IRMの基本音列パターン例

この三つ組パターンを単位として、それで表された旋律の解決・未解決などの関係を分類・分析し、楽曲の構造的特徴を記述しようというのがナムアのIRM理論の骨子である。ただし、実際には細かい各論で論じられるため、全体像を簡単にまとめるのは難しい。

IRMは暗意とその(肯定的・否定的)実現に基づくため、本質的に動的な構造記述理論である。これはGTTMにおける緊張・弛緩関係と共通する部分はあるが、GTTMが最終的には階層構造に帰着させるのに対し、暗意は様々な箇所が発生し、その実現・解決も多義的なため、それらの関係として生成される構造は多次元的なネットワーク構造になる。その分、形式的解析は困難になるが、人間が実際に音楽から聴き取る内容を、的確に表現し得るという面もある。そのため、GTTMよりは認知理論的な性格が強い。

1-3-3 計算機上への実装

GTTM を計算機上へに実装しようという試みは数多くある。しかし本にある解析例を忠実に再現し得るようなシステムは存在しない。特に延長的簡約まで扱った例はなく、タイムスペース簡約を扱う浜中らのシステム¹⁰⁾は最も先進的な研究例といえる。

実装が難しい理由の一つは、選好ルール間に競合が生じたとき、どう解消すべきかが明確でない点にある。選好ルールそのものが競合を前提とした枠組みだが、GTTM も明確な規定は示していない。また類似したパターンを検出や和声解析とそれに基づく緊張・弛緩関係の判断などはルールには組み込まれているものの、具体的な定式化は(意図的に)避けられている。これら高次の処理を計算機上で実現することは、それ自身が難しい研究テーマである。

一方、IRM については実装の試みそのものがほとんど存在しない。これは理論自体の魅力にもかかわらず、ナムアの記述の膨大さ、難解さもあいまって、何がどう実現できればいいのか自体が判断しづらいのが大きな理由だろう。

一般に何らかの理論を計算機上へに実装するというのは、実現された場合に生ずる様々な用途に対する有用性ととも、理論そのものがプログラムという形式的・機械的処理に適用し得るほど精密に定式化されているか、理論に抜けや誤りがないかを検証する手段になるという利点もある。その一方で、実装に当たっては様々な簡略化や仮定の導入が行われるが、それが理論の本質を変節させたりしない注意が必要となる。

■参考文献

- 1) L.B. Meyer, "Emotion and Meaning in Music," Univ. of Chicago Press, 1956.
- 2) L.B. Meyer, "Music, the Arts, and Ideas," Univ. of Chicago Press, 1967.
- 3) L.B. Meyer, "Explaining Music: Essays and Explorations.," Univ. of California Press, 1973.
- 4) F. Lerdahl and R. Jackendoff, "A Generative Theory of Tonal Music," MIT Press, 1983.
- 5) H. Schenker (E. Oster (trans.)), "Free Composition," Longman, 1979. (原著: "Der Freie Satz," Universal Edition, 1935/1956 (Rev. Ed.))
- 6) 竹内好宏, "音楽の構造解析とその応用," 長嶋・橋本・平賀・平田(編), "コンピュータと音楽の世界"(bit 別冊), pp.224-240, 共立出版, 1998.
- 7) E. Narmour, "Beyond Schenkerism - The Need for Alternatives in Music Analysis," Univ. of Chicago Press, 1977.
- 8) E. Narmour, "The Analysis and Cognition of Basic Melodic Structures - The Implication-Realization Model," Univ. of Chicago Press, 1990.
- 9) E. Narmour, "The Analysis and Cognition of Melodic Complexity - The Implication-Realization Model," Univ. of Chicago Press, 1992.
- 10) M. Hamanaka, K. Hirata and S. Tojo, "Implementing 'A Generative Theory of Tonal Music'," Journal of New Music Research, vol.35, no.4, pp.249-277, 2007.

■2群-9編-1章

1-4 音楽記述（表現）

（執筆者：平田圭二）[2011年6月受領]

音楽とは、音を手段としそれらの時間的な動きや変化によって表現する芸術（時間芸術）であり、音楽の記号表現は記号論的な側面と計算機上での操作の側面をもつ。本項前半では記号論的な側面について述べ、後半で計算機上での操作に関連した事項について述べる。

1-4-1 記述力と簡潔さ

楽曲中に現れる時間的な動きや変化を伴う複数の音は、西洋音楽において3要素と考えられている**旋律**、**和声**、**リズム**という概念を作り出している。音楽の記述とは、ここに挙げた、時間、音の動きや変化、旋律、和声、リズムといった概念や楽曲を構成する部品（音楽イベント）を記号の列として記述することであり、これらを記号列に対応づけるときに考慮すべきは、記述力と簡潔さのトレードオフである*。

記述力は、ある概念や音楽イベントをどれだけ具体的に記述できるかの指標である。例えば和音の記述において、抽象度の高い場合から低い場合へと、和音名 C_7 、音の集合 $\{C E G B b\}$ 、五線譜上に記された和音、と並べることができる。一般に、抽象的な表現ほど情報量が少ない。あるいは、その記述からもとの楽曲がどの程度復元できるかで記述力を測ることもできる。例えば、リズムのもつグルーブ感（groove、ノリ）に関しては、リズムパターンに含まれる各音の発音タイミングを物理的に記述できた方が、できないより記述力が高い。なぜなら、前者ならばもとの音響信号としての楽曲が復元できるからである。

簡潔さは書くコストと読み取りコストの積と言い換えてもよく、人にとっての可読性や、機械にとっての操作性、構文解析の容易さに関係する。例えば、旋律に含まれる各音の音の高さを記述する方法として、調とその音階上の階名、ピッチ†、旋律中の直前の音との音程‡などがある。これらの内、音程による記述が最も簡潔であり、調とその音階上の階名による記述が最も複雑であろう。

一般に、記述力と簡潔さを両立させることは難しい。例えば、4声からなる旋律に対して、記述力の高い例として、声部ごとの音の動きに着目した記述 a と、同時に鳴っている4声に着目した記述 b の両方を含むような記述を考える。まず a では、まず声部ごとの記述が作成され、得られた四つの記述がグルーピングされる。次に b では、まず同時に鳴っている4声グルーピングされ、それらが時間順に並べられる。この a、b 両方を含むような記述を書くコストは、a のみあるいは b のみより大きく、簡潔ではない。

音楽記述の方法を設計したり選択したりする場合は、その目的に関して、記述力と簡潔さのいずれに優先度を置くのかよく吟味すべきである。

* 知識表現の分野では、記述力（descriptive power）対簡潔さ（simplicity）のほかに、expressiveness 対 tractability, generality 対 efficiency という場合もある。

† 一般に、ピッチ（pitch）はオクターブ位置と音名（C, D, E, ..., ハ, ニ, ホ, ...）の組である。音の高さを表すパラメータはピッチのほかに物理的な周波数を単位とする「音高」がある。しかし実際には音高をピッチと同義に用いる場合もある。

‡ 音程には、時間的に隣接する2音に関する音の高さの差（水平の音程）と、同時に鳴る2音に関する音の高さの差（垂直の音程）の意味がある。

1-4-2 グループと構造

簡潔な記述を実現するには、音楽がもつ時間、音の動きや変化、旋律、和声、リズムといった概念を用いるのが効果的である。それは以下のような理由による。音のピッチ方向のゲシュタルトの知覚と時間方向のゲシュタルトの知覚が、一つひとつの音を分節させると同時に、旋律、和声、リズムといった音楽的な音楽イベントを生じさせる。ここでゲシュタルトの知覚とは、近接した音どうし、類同した音どうし、閉合した音どうし、連続性のある音どうしはグループを作るという知覚現象であり、そうやって生じたグループ間にも更にゲシュタルトが知覚される。その結果、グループは階層構造やネットワーク構造を作る。これらグループ、グループ間に作られる関係、そうやってできあがった構造が楽曲の知覚、すなわち旋律、和声、リズムの知覚に対応していると考えられている。そして、旋律、和声、リズムという構造に沿って楽曲を記述すれば、人が音楽を記述するときに自然と出てくる表現、例えば「旋律中の次の音」、「その音を含む和音全体の音」、「リズムパターンの1周期」などがその構造の中に明示的に存在しているため、簡潔に記述できるのである。

記述力を上げるためには、単純には、構造の中にできる限り多くの音楽的な概念を明示的に導入すればよい。しかし、過剰に導入すると、目的に関して不要な記述が増えたり、記述が冗長になったり、記述の一意性が失われる場合があるので注意が必要である。例えば、前述の4声からなる旋律を記述する方法ではaとbの双方を含むので目的によっては冗長な場合がある。そのようなときは、aまたはbのいずれかで十分である。別の例として、グループ感のあるリズムを表現するために、逸脱[§]という概念を導入し、各音の発音タイミングをミリ秒単位で前後させる記述法が利用できるようになったとしよう。この逸脱の記述法を濫用し、和音の各音の発音タイミングを数ミリ秒の値ではなく数秒の値で逸脱させると、同じに聞こえる旋律を実現する意味のない記述が複数通り存在してしまう。一方で、結果として同じ音の並びや音響信号を、書き手の意図を反映して複数通りの記述で表現し分けることができる(記述力が高い)のは望ましい。優れた音楽記述法とは、書き手の意図を反映した複数通りの意味のある記述を許しながら、濫用をできるだけ少なく抑えられるような記述体系である。

1-4-3 五線譜という音楽記述法

五線譜(Common Western Music Notation, CMN)では、音符として表現される一つの音がその音楽イベントに相当する。個々の音を記述するパラメータは音の長さ(音価^{**})と音の高さ(ピッチ)である。音の長さと言音の高さという二つの特徴あるいは属性が用いられて、音という一つのゲシュタルト知覚が生じた、あるいは音響信号を分節して音を知覚したと考えられる。

五線譜では、一つひとつの音を縦軸がピッチ、横軸が発音時刻に対応する拍であるような2次元平面の上に配置する。ピッチの高いものほど五線譜の上の方に配置され、発音順に左から右に並べられる。ここで注意すべきは、時間の近い音どうしほど、また音の高さの近い

[§] 各音楽イベントの発音タイミングや音高を、楽譜を機械的に演奏したときのそれら(deadpan演奏とも呼ばれる)から意図的にズラすこと。

^{**} 一般に音価の単位は拍である(知覚量)。音の長さを表すパラメータは音価のほか単位が秒やミリ秒などの物理的な時間である音長(length)がある(物理量)。しかし実際には音価と音長の使い分けは曖昧である。

音どうしほど、2次元平面上で近い位置に配置される点である。つまり、五線譜は音楽的な距離を視覚的な距離に対応づけることで、音楽中で生ずるゲシュタルト知覚を簡潔に表現しているのである。

一方で、音を2次元平面に配置したためにピッチと拍から算出されるような距離/近さが生ずる。例えば、半音と1拍を単位長とすると、3拍後の4半音異なる音との距離は5となるが(各辺の長さが3, 4, 5であるような直角三角形より)、この距離5は、同じ音高で5拍後の音どうしや、同時刻で5半音離れた音どうしと同じ近さをもっているのかという音楽的な問題が生ずる。このように、音楽記述法では、音楽イベントを表現する記号の並べ方によって暗黙的に生じてしまう情報が意図したものであるか否かに注意しなければならない。

五線譜には更に、音楽的な意味を付加したり補助したり、可読性を上げるための記号や書法が多数存在する(例えば、音部記号、調号、拍子記号、小節線、音、連桁、臨時記号(b, #, ♯), スラー、演奏記号^{††}など)。これらは五線譜の記述力を高めることに貢献している。

1-4-4 計算機上での操作に関する事項

音楽記述の問題は、音楽を記号の列として記述するという課題に加えて、記述されたデータを実際に計算機上で使用することに関する現実的な課題も含んでいる。具体的には以下のような留意点が挙げられる。(1) 利用目的(作曲、分析、データベース化など)に関して、要請された機能を満足しているか(特殊奏法の記述や音響信号データとの連携)、仕様が過剰になっていないか、(2) 世の中に広く普及し、ツール群やフレームワークが完備されているか、(3) 他の記述法との相互運用性は高いか、(4) 仕様がオープンになっており関連ドキュメントが充実しているか、(5) 長期間に渡り維持、改良、拡張、修正される見通しはあるか、(6) 現時点でどの程度の量のデータやアーカイブが入手可能か、(7) 自分の用途はその記述データの利用条件を満足しているか(特に法律的な事項について)。

1-4-5 代表的な音楽記述法

(1) 標準MIDIファイル形式

MIDI (Music Instrument Digital Interface) は演奏記述のためのデファクト標準規格である¹⁾。MIDIにおける音楽イベントはNote on (鍵盤を押下するアクション)、Program change (音色切替えのアクション) などシンセサイザなどの電子楽器の鍵盤やフロントパネル上で可能な動作であり、MIDIコマンドとも呼ばれる。MIDIコマンドを受信した電子楽器は、受信直後にできるだけ遅延なくMIDIコマンドを実行することで演奏を再現することができる。MIDIコマンドとその実行タイミングを指示すれば、演奏を記述することができる。この形式は標準MIDIファイル形式(Standard MIDI File (SMF) Format) と呼ばれる。

(2) MusicXML

MusicXMLはXMLベースの五線譜向き楽譜記述言語である^{2,3)}。楽譜記述とは、楽曲や演奏そのものを記述するのではなく、紙に描かれることの多い楽譜を記号の列として記述する言語である。MusicXMLは五線譜上に描かれるすべての記号(上述の音部記号、調号、拍子記号、小節線、音、連桁など)と暗黙的なパラメータ(時間軸の最小単位など)に対応する

^{††} 演奏時のテンポ、強弱、演奏法、発想などを指示するため楽譜に書き加えられた記号や文字。

XML 要素 (element) を含んでいる。2 次元平面に描かれた五線譜を 1 次元の記号列に変換するために 2 次元平面をスキャンする二つの方法が提供されている。一つ目は、まずあるパートに着目しそのパートを時間軸方向をスキャンし、パートごとの記号列を得る方法 (partwise) である。もう一つは、同時刻における音高方向を先にスキャンし時刻ごとの記号列を得る方法 (timewise) である。MusicXML に関連するツール群やほかの記述法との相互運用性などについては <http://www.recordare.com/good/cm12.html> に詳しい。

(3) DARMS, MML

MIDI や MusicXML が誕生する以前より、計算機に楽譜情報を簡易に入力するための言語が存在している。最も古いものの一つは Erickson が 1975 年に提案した DARMS (Digital Alternate Representation of Music Scores) であり、現在でもその簡易性ゆえに使われているものとして MML (Music Macro Language) などがある。これらはいずれも、楽譜を timewise に記号化する。

(4) その他

楽譜記述言語の一つである Humdrum⁴⁾は、五線譜で記述された楽曲に関して、ある特徴をもつ旋律を検索したり、統計的な情報を算出したりするような用途 (楽曲分析) にも有用である。人がテキストとしても読みやすいように記述自体が工夫されている。

XML の特徴を活かし音楽がもつ階層的な構造を反映し、楽譜情報と音響信号を同時に記述できるような形式として提案されたものとして IEEE1599⁵⁾や CrestMuseXML⁶⁾がある。IEEE1599 は楽譜、演奏、歌詞、音響信号など種々のメディアを統合し同期させることを目標にゼロから新しく設計された規格である。対して CrestMuseXML は、デファクト標準となりつつある MusicXML を中心に据えて、XPath/XPointer で必要な情報を付加し (MusicXML には一切変更を加えずに) 他メディアを統合し同期させる規格である。

■参考文献

- 1) MIDI Manufactures Association, <http://www.midi.org/>
- 2) Michael Good, "Representing Music Using XML," In Proceedings of ISMIR 2000.
- 3) Recordare, <http://www.musicxml.org/xml.html>
- 4) CCARH Humdrum Portal, <http://humdrum.ccarh.org/>
- 5) Denis Baggi and Goffredo Haus, "IEEE 1599: Music Encoding and Interaction," IEEE Computer, vol.42, no.3, pp.84-87, 2009.
- 6) 北原鉄朗, CrestMuseXML Development Project, <http://www.crestmuse.jp/cm/x/>

■2群-9編-1章

1-5 スタイル・ジャンル

(執筆者：平田圭二) [2011年6月受領]

1-5-1 スタイル・ジャンルの意味

多数の人が所属するコミュニティに多数の楽曲が存在するとスタイルやジャンルが発生する。このとき、1曲だけではスタイルやジャンルとは呼ばれないとすると、同じような楽曲が何曲くらい集まったらスタイルやジャンルと呼ばれるようになるのだろうか。また、スタイルやジャンルは一般に階層構造をなしているが^{††}、それは、あまりに幅の広い楽曲群を一つのスタイルやジャンルにくくるのは不自然だからである。では、スタイルやジャンルを定めている境界線のある曲が越えているかどうかはどのようにして決まるのだろうか。つまり、スタイル・ジャンルは、個々の楽曲間に共通する要素が人々に認識されるようになることと(前者の共通性の認識)、個々の楽曲を比べたときお互いどこを特異な部分と認識されるようになること(後者の特異性の認識)の2面から決まる。

音楽辞典¹⁾によれば、音楽学におけるスタイル(様式)とは、音楽作品のもつ具体的な特性に関して形式的特徴として現れる類型のこととある。類型の例としては、楽式(楽曲の形式)、表現材料や題材(神話、事件)、楽器(ピアノ、ヴァイオリン)、時代(バロック、現代)、個人の素質、人格、技巧(バッハ、モーツァルト)、民族(スペイン、中華、アフリカ)、流派(ウィーン、ヴェネチア)などが挙げられる。ジャンルも同様に、何らかの音楽的な分類を意味する。これらの類型は、共通性の認識と特異性の認識が互いに影響を及ぼしあいながら、時間をかけてコミュニティ内で到達した合意の一部であり、共通性の認識と特異性の認識のトレードオフといてよいだろう²⁾。

しかし、音楽情報処理におけるスタイル・ジャンルの意味は、音楽用語のそれとは大きく異なっている³⁾。音楽情報処理におけるスタイル・ジャンルとは、楽曲群を機械的に分類するために決められた名前(ジャンル名、スタイル名^{§§})の集まりのことである。音楽用語としてのスタイル・ジャンルのように音楽的な根拠をもってその名前が付けられる場合もあれば、音楽的な理由とは無関係な理由でクラスタリングを行い名前が付けられる場合もある(例えばビジュアル系など)。

1-5-2 二つのアプローチ

自動スタイル・ジャンル分類のアプローチは大きく、音響信号に基づく方式(content-based)とメタデータに基づく方式(metadata-based)に分けられる。これは、楽曲間の類似度あるいはアーティスト間の類似度を算出するのに、音響信号の音響特徴量を使う方法とメタデータを使う方法が有効だからである。前者の音響信号に基づく自動スタイル・ジャンル分類に関する優れたサーベイとして文献4)がある。後者には、様々な手段で獲得されたメタデータに基づく方法⁵⁾や、Web上での共起単語を用いた協調フィルタリングに基づく方法⁶⁾などがあ

^{††} 場合によっては、一つの楽曲が複数のジャンルに属することもあり、厳密には階層構造ではない。

^{§§} ジャンル名の例にはジャズ、ポップ、ロック、クラシックなどがある。スタイル名は一般に、ジャンルより更に細分化された分類を指す場合が多い。例えば、ロックは更にハードロック、ヘビメタル、プログレ、グラム、パンクなどサブカテゴリに分かれている。

る。

一般に、楽曲群をいくつかのクラスタに分割する属性や尺度は、上述の共通性の認識や特異性の認識に対応して、音響的なものからテキストによるものまで多種類考えられ、それらは通常異なるクラスタリング結果をもたらす。例えば和声進行という尺度による楽曲群の分割と、テンポという尺度による楽曲群の分割は異なる。自動スタイル・ジャンル分類の研究とは、互いに異なるクラスタリング結果をもたらすような尺度を適切に組み合わせ、目標とするスタイル・ジャンル分類を精度高く実現する課題と換言できる。更に、付与されるスタイル・ジャンル名が2個以上の場合もある。

音楽の専門知識のない人でもある程度安定してスタイル・ジャンル分類できるということから、それほど高度な音楽知識を仮定せず、人が普通にもっているような識別能力を実現すれば、ある程度まで自動的にスタイル・ジャンル分類できるという仮説が支持されている。実際、音響信号から様々な音響特徴量を取り出して特徴ベクトル間の距離を算出するアプローチでも用途を限定すれば実用的に十分な場合がある。しかし、人でも識別の間違いが起きるようなスタイル、例えば時代、民族、流派を音響特徴量だけで識別するのは難しい。このような場合は、上述の二つのアプローチを混合したり⁷⁾、高度な音楽知識を利用することが効果的である。

1-5-3 研究事例

スタイル・ジャンル分類に利用される音響の特徴量の多くは、人が楽曲を聴取するときに注意を向けている特徴や部分はどこかという内省に基づいて見出されたものである。音楽の3要素は旋律、和声、リズムといわれるが、この内、旋律と和声に関する特徴はスペクトログラムのテクスチャとして出現する音響の特徴に関連づけられ、リズムに関する特徴は時間方向の周期的パターンとテンポに関連づけられることが多い。着目した特徴に関する情報を楽曲全体に渡って収集する方式(bag-of-frames)と、時間順序に沿って収集する方式がある。音色、テンポ、旋律の各マクロ情報を用いた研究の最初期のものに文献8)がある。マイクロ情報として、リズム⁹⁾、低音旋律¹⁰⁾、和声認識^{11, 12)}などを用いる研究がある。和声認識をした後、時間軸に沿って和音進行どうしの類似度を比較するのにDynamic Programming (DP)やHidden Markov Model (HMM)が用いられることが多い。次に、所与のスタイル・ジャンル名の集合と特徴量ベクトルの集合を対応づけるため、パターン認識を行ったり、統計的なモデルを仮定してそのパラメータを機械学習する⁴⁾。一方、記号処理レベルで、和声に基づくジャンル分類を行う研究もある¹³⁾。

スタイル・ジャンル分類に利用される重要な音響の特徴量や記号の特徴量に関しては、毎年MIREX (Music Information Retrieval Evaluation eXchange) というコンテストが開催されている¹⁴⁾。音楽情報検索 (MIR) は記号処理、信号処理、テキスト処理など多分野にまたがる研究領域なので、最新のMIRアルゴリズムやMIRシステムの性能比較は容易ではない。そこで、Audio Genre Classification, Audio Artist Identification, Symbolic Melodic Similarityなどの要素技術に細分化し、与えられたベンチマークに関して性能を直接比較する。第1回目のMIREXコンテストはISMIR 2005 (London) 会期中に開催された。

■参考文献

- 1) 新音楽辞典楽語, 音楽之友社, 1977.
- 2) Cory McKay and Ichiro Fujinaga, "Musical genre classification: Is it worth pursuing and how can it be improved?," in Proc. of ISMIR 2006.
- 3) Jean-Julien Aucouturier and Francois Pachet, "Representing Musical Genre: A State of the Art," Journal of New Music Research, vol.32, no.1, pp.83-93, 2003.
- 4) George Tzanetakis, "音響ベースの音楽信号分類," 情報処理, vol.50, no.8, pp.746-750, 角尾衣未留 (訳), 2008.
- 5) Dan Ellis, Brian Whitman, Adam Berenzweig, and Steve Lawrence, "The quest for ground truth in musical artist similarity," in Proc. of ISMIR 2002.
- 6) Peter Knees, Elias Pampalk, Gerhard Widmer, "Artist Classification With Web-Based Data," In Proc. ISMIR 2004.
- 7) Brian Whitman and Paris Smaragdis, "Combining Musical and Cultural Features for Intelligent Style Detection," in Proc. of ISMIR 2002.
- 8) George Tzanetakis and Perry Cook, "Musical Genre Classification of Audio Signals," IEEE Transactions on Speech and Audio Processing, vol.10, no.5, pp.293-302, 2002.
- 9) Jouni Paulus and Anssi Klapuri, "Measuring the similarity of rhythmic patterns," in Proc. of ISMIR 2002.
- 10) Yusuke Tsuchihashi, Tetsuro Kitahara and Haruhiro Katayose, "Using Bass-line Features for Contentbased MIR," in Proc. of ISMIR 2008, pp.620-625.
- 11) Alex Sheh and Dan Ellis, "Chord Segmentation and Recognition Using EM-Trained Hidden Markov Models," in Proc. of ISMIR 2003.
- 12) Takuya Yoshioka, Tetsuro Kitahara, Kazunori Komatani, Tetsuya Ogata and Hiroshi G. Okuno, "Automatic Chord Transcription With Concurrent Recognition of Chord Symbols and Boundaries," in Proc. Of ISMIR 2004.
- 13) Amelie Anglade, Rafael Ramirez and Simon Dixon, "Genre Classification Using Harmony Rules Induced from Automatic Chord Transcriptions," in Proc. of ISMIR 2009.
- 14) [http://www.music-ir.org/mirex/2005/index.php/Main Page](http://www.music-ir.org/mirex/2005/index.php/Main_Page)

■2群-9編-1章

1-6 音楽情報検索

(執筆著: 後藤真孝) [2011年6月 受領]

「音楽情報検索」(MIR: Music Information Retrieval) は、音楽の検索、推薦、分類、管理、閲覧などに幅広い観点から取り組む研究テーマ群の総称である。2000年代の音楽情報処理分野を代表する研究トピックであり、国際会議 ISMIR (2000~2001年は International Symposium on Music Information Retrieval, 2002~2008年は International Conference on Music Information Retrieval, 2009年以降は International Society for Music Information Retrieval Conference) が2000年以降毎年開催されるなど、活発に研究されている^{1,2,3)}。ISMIRの参加者は当初100名に満たなかったが、近年は200名を越え、日本で開催された ISMIR 2009では、過去最高の29か国308名の参加があり、関心が高まっている。ISMIR以外にも、様々な会議、論文誌などで音楽情報検索の研究成果が報告されており、広義に音楽情報検索に関連した研究をした博士学位取得者は、2004年以降、少なくとも76名いる⁴⁾など、学術的にも重要な分野である。

音楽情報検索、つまり音楽を対象に情報検索するという考え方は、曲を一単位として大量の楽曲の集合を計算機上で扱う問題意識が出てきて、初めて誕生した。1990年代のハミング検索 (QBH: Query By Humming) の研究がその起源である。音楽情報検索が扱う問題は、2000年代を通じて拡大しつつあり、ユーザがメロディをハミングしたり歌ったりして探すハミング検索以外にも、ユーザが曲名を知りたい楽曲の断片を与える検索、ユーザが気に入った楽曲に曲調や声質が類似した楽曲を探す検索などが研究開発された。これらは楽曲内容の分析・理解技術に基づいており、内容に基づく音楽情報検索 (Content-based Music Information Retrieval) と呼ばれる⁵⁾。

一方、最も単純な書誌情報 (CDDBによる曲名やアーティスト名など) に基づくテキスト検索は、古くから商用音楽サービスなどで一般的であった。近年では、それに加えて、ユーザが自分たちで楽曲やアーティストに対してタグ (ラベル) を付与するソーシャルタギングも普及し、タグ検索やタグクラウドなども重要な楽曲検索手段となっている。これらのタグやラベルは、音楽の内容を人間が解釈して表現した **アノテーション (Annotation)** あるいは **メタデータ (Metadata)** ととらえることができる。近年は、ゲーム形式でユーザに自発的にアノテーション付与を促す Web サイトも様々なアプローチで試みられ、集積されたアノテーションを研究コミュニティで共有する動きも始まっている。より詳しい解説は、[1-8 メタデータ・アノテーション (梶)] 節で述べられているため、本節では割愛する。

検索のように明示的に検索キーを与えなくても、ユーザの好み (音楽の嗜好) を分析して楽曲やアーティスト、ジャンルなどを薦める **音楽推薦 (Music Recommendation)** も重要な技術である。漠然と音楽が聴きたい状況では、能動的に何かを探すための検索キーを与えることが難しいことは多く、受動的に楽曲を自動的に選択して再生してくれる方が便利だからである。そうした楽曲の再生順を指定する **プレイリスト (Playlist)** を自動生成するプレイリスト生成技術も、音楽推薦の一つの形態といえる。音楽推薦を実現するうえでは、協調フィルタリング (Collaborative filtering) 技術や内容に基づくフィルタリング (Content-based filtering) 技術、あるいはそれらの組合せが代表的であるが、より詳しい解説は、[2-5 音楽推薦・プレイリスト生成 (吉井)] 節で述べられているため、本節では割愛する。

以下、音楽情報検索分野で比較的初期の頃から研究されていたハミング検索、断片を含む楽曲の検索、楽曲間の類似度に基づく検索を説明する。

1-6-1 ハミング検索 (QBH: Query by Humming)

聞いたことのある曲を「ラララー」などのように口ずさむと、その曲名を検索できる方法である。つまりメロディの歌唱やハミングを検索キーとして、そのメロディをもつ楽曲を検索する。多くの場合、入力したメロディと、データベース中の楽曲のメロディとの類似度順に候補を出すこともできる。

1991年の蔭山ら^{6,7,8)}による研究、1995年のGhiasら⁹⁾による研究、1997年の貝塚ら¹⁰⁾、園田ら^{11,12,13)}による研究が初期を代表する研究であり、日本で当初は活発に研究されていた。ハミング検索に関しては、園田らの技術^{11,12,13)}が2004年11月に世界で初めて実用化されて、携帯電話で利用可能となった。その後も、国内外で様々なハミング検索サービスが実用化され、提供されている。

検索キーについては、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などが課題となる。具体的な手法は、データベースとして、メロディのみ⁶⁻¹⁵⁾、楽曲全体の標準MIDIファイル (SMF)^{16,17,18)}、楽曲全体の音響信号^{19,20)}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、より実現が困難となる。また、検索キーがハミングでなく歌詞付きの歌唱の場合には、メロディだけでなく、歌詞も手がかりとして併用して検索する方法²¹⁾も提案されている。

1-6-2 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、携帯電話など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。つまり、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。

高速化とともに、雑音や伝送路の歪みなど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²²⁾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²³⁾などが提案されている。

断片を含む楽曲の検索に関しては、Wangらの技術²³⁾が2002年8月に世界で初めて実用化されて、携帯電話で利用可能となった。その後も、国内外で様々な同様のサービスが実用化され、提供されている。

1-6-3 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す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 (Similarity) を、楽曲中の音色 (パワースペクトル形状)^{24,25)}、リズム²⁵⁻²⁸⁾、変調スペクトル²⁹⁾、歌声³⁰⁾など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)^{25, 28~31)}なども研究されている。

楽曲どうしあるいはアーティストどうし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、音楽情報検索における本質的な技術課題の一つとして現在でも活発に研究がなされている。音楽音響信号理解技術の進展とともに、今後の更なる研究が期待される。

1-6-4 おわりに

本節では音楽情報検索の研究トピックを、狭義の「検索」に関連する技術を中心に紹介したが、国際会議 ISMIR で発表されているような広義の音楽情報検索に関連した技術は幅広く、例えば、音楽分析・理解 (Music Analysis / Understanding)、音源分離・採譜 (Source Separation / Transcription)、対応づけ・同期 (Alignment / Synchronization)、類似度 (Similarity)、分類 (Classification, Categorization)、可視化 (Visualization)、コンテキスト (Context)、ユーザモデリング (User Modeling)、ユーザインタフェース (User Interface)、データベース・コーパス (Database / Corpus) などが挙げられる。また、技術課題を解決するうえで、音響信号処理、記号処理、機械学習、Web 技術、音楽学、図書館情報学などが求められ、学際的な研究トピックになっている。

Web 上でアクセス可能な音楽・動画が増加し続け、1000 万曲以上が定額制聴き放題のオンライン音楽配信サービスも登場するなど、好きな曲を好きなときに好きな場所で好きなだけ聴ける時代になった。音楽情報検索・推薦等に関する需要は今後もますます高くなることが予想され、更なる技術的な進展が望まれる。

■参考文献

- 1) ISMIR - The International Society for Music Information Retrieval, <http://www.ismir.net/>
- 2) J.H. Lee, M.C. Jones, and J.S. Downie., "An analysis of ISMIR proceedings: Patterns of authorship, topic, and citation.," Proceedings of ISMIR 2009, pp.57-62, 2009.
- 3) M. Grachten, M. Schedl, T. Pohle, and G. Widmer, "The ISMIR Cloud: A decade of ISMIR conferences at your fingertips.," Proceedings of ISMIR 2009, pp.63-68, 2009.
- 4) PhD Theses and Doctoral Dissertations Related to Music Information Retrieval, <http://www.pampalk.at/mir-phds/>
- 5) M. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney. Content-based music information retrieval: Current directions and future challenges, Proceedings of the IEEE, Vol.96, No.4, pp.668-696 (2008).
- 6) 蔭山哲也, 島津秀雄, 高島洋典. メロディ検索- ハミングで音楽DBを検索する, 情報処理学会第 43 回全国大会, 4-149 (1991).
- 7) T. Kageyama, K. Mochizuki, and Y. Takashima, "Melody retrieval with humming.," Proceedings of ICMC 1993, pp.349-351, 1993.
- 8) 蔭山哲也, 高島洋典, "ハミング歌唱を手掛りとするメロディ検索," 電子情報通信学会論文誌 D-II, vol. J77-D-II, no.8, pp.1543-1551, 1994.
- 9) A. Ghias, J. Logan, D. Chamberlin, and B.C. Smith, "Query by humming: Musical information retrieval in an audio database.," Proceedings of ACM Multimedia 95, pp.231-236, 1995.
- 10) 貝塚智憲, 後藤真孝, 村岡洋一, "歌声の旋律情報と歌詞情報をキーとした曲検索システム," 情報処理学会第 54 回全国大会, vol. 7J-06, 1997.
- 11) 園田智也, 後藤真孝, 村岡洋一, "歌声による曲検索システム-音程・音長情報の分布を利用した閾値の決定法-, " 情報処理学会第 55 回全国大会, vol. 1J-06, 1997.
- 12) T. Sonoda, M. Goto, and Y. Muraoka, "A WWW-based melody retrieval system.," Proceedings of ICMC 1998, pp.349-352, 1998.
- 13) 園田智也, 後藤真孝, 村岡洋一, "WWW 上での歌声による曲検索システム," 電子情報通信学会論文

誌 D-II, vol.J82-D-II, no.4, pp.721-731, 1999.

- 14) S. Pauws, "CubyHum: A fully operational query by humming system," Proceedings of ISMIR 2002, pp.187-196, 2002.
- 15) T. Sonoda, T. Ikenaga, K. Shimizu, and Y. Muraoka, "The design method of a melody retrieval system on parallelized computers," Proceedings of WEDELMUSIC 2002, pp.66-73, 2002.
- 16) J. Shifrin, B. Pardo, C. Meek, and W. Birmingham, "HMM-based musical query retrieval," Proceedings of JCDL 2002, pp.295-300, 2002.
- 17) N. Hu and R.B. Dannenberg, "A comparison of melodic database retrieval techniques using sung queries.," Proceedings of JCDL 2002, pp.301-307, 2002.
- 18) R.B. Dannenberg, W.P. Birmingham, G. Tzanetakis, C. Meek, N. Hu, and B. Pardo, "The MUSART testbed for query-by-humming evaluation.," Proceedings of ISMIR 2003, pp.41-47, 2003.
- 19) 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.," Proceedings of ISMIR 2001, pp.211-218, 2001.
- 20) J. Song, S.Y. Bae, and K. Yoon, "Mid-level music melody representation of polyphonic audio for query by humming system.," Proceedings of ISMIR 2002, pp.133-139, 2002.
- 21) M. Suzuki, T. Hosoya, A. Ito, and S. Makino, "Music information retrieval from a singing voice using lyrics and melody information," EURASIP Journal on Applied Signal Processing, vol.2007, no.1, 2007.
- 22) K. Kashino, T. Kurozumi, and H. Murase, "A quick search method for audio and video signals based on histogram pruning.," IEEE Transactions on Multimedia, vol.5, no.3, pp.348-357, 2003.
- 23) A. Wang, "An industrial-strength audio search algorithm.," Proceedings of ISMIR 2003, pp.7-13, 2003.
- 24) J.-J. Aucouturier and F. Pachet, "Music similarity measures: What's the use? ," Proceedings of ISMIR 2002, pp.157-163, 2002.
- 25) G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," IEEE Transactions on Speech and Audio Processing, vol.10, no.5, pp.293-302, 2002.
- 26) J. Paulus and A. Klapuri, "Measuring the similarity of rhythmic patterns.," Proceedings of ISMIR 2002, pp.150-156, 2002.
- 27) J. Foote, M. Cooper, and U. Nam, "Audio retrieval by rhythmic similarity.," Proceedings of ISMIR 2002, pp.265-266, 2002.
- 28) S. Dixon, E. Pampalk, and G. Widmer, "Classification of dance music by periodicity patterns.," Proceedings of ISMIR 2003, pp.159-165, 2003.
- 29) M.F. McKinney and J. Breebaart, "Features for audio and music classification.," Proceedings of ISMIR 2003, pp.151-158, 2003.
- 30) 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.," Proceedings of ISMIR 2003, pp.167-173, 2003.
- 31) B. Whitman and P. Smaragdus, "Combining musical and cultural features for intelligent style detection.," Proceedings of ISMIR 2002, pp.47-52, 2002.

■2群-9編-1章

1-7 音楽データベース

(執筆者：後藤真孝) [2011年6月 受領]

音楽データベース (音楽DB, music database) とは、音楽に関連したデータを集積して再利用を容易にしたデータの集合である。音楽コーパス (music corpus) と呼ばれることもある。データとしては、楽曲、書誌情報 (アーティスト名や曲名など)、楽器音、歌声、歌詞、演奏表情などの様々な対象が考えられる。それを記録する形態は、音響信号ファイル (RIFF WAVE フォーマットのファイルなど)、MIDIファイル (標準MIDIファイル (Standard MIDI File) など)、XMLファイル (RecordareのMusicXMLなど)、テキストファイルなどから、データごとに適切なかたちが選択される。このように広義には、音楽データが集積すれば音楽データベースとみなせるため、企業などが所有する非公開のものまで含めれば、既に膨大な音楽データベースが世の中には存在する。しかし、ここでは研究目的での利用という観点から、デジタルデータで、かつ研究で実際に利用可能な音楽データベースを中心に述べる。

1-7-1 学術研究における著作権処理済み音楽データベースの重要性

音楽に関連した研究を進めるうえで、自由に利用できる音楽データベースは重要な役割を果たす。例えば、様々な研究者が、個々のシステムや手法を適切に比較・評価するためには、ある楽曲集をベンチマークとして共通に利用するとよい。統計的手法や学習手法を活用した研究でも、一定規模のデータベースが必要となる。更に、研究成果を学会などで対外発表する際にも、自由に音楽を利用できると便利なおことが多い。以上は主に音楽情報処理の観点からの利用法だが、音楽心理学、音声言語情報処理などの観点からも、共通利用・学術利用の自由が確保された研究用の音楽データベースは有用となる。

しかし、普段私たちが耳にする市販の楽曲 (音楽コンパクトディスクや音楽配信で商業的に提供される楽曲) の多くは、著作財産権、著作人格権、著作隣接権などの制約のため、世界中の研究者が自由に使うのは困難である。例えば、研究成果の発表で、楽曲をサンプルとして用いたシステムのデモンストレーションを Web に掲載したり、学会発行の CD-ROM に収録したりする際には、著作財産権が関係する。著作財産権の保護期間が終了したクラシック音楽などを用いたとしても、近年の録音物は著作隣接権 (演奏家や歌手、レコード製作者などのための権利) で保護されている。したがって、基本的にこうした問題を克服するためには、新規に作曲、作詞するか、クラシック音楽などの著作財産権の保護期間が終了した既存曲を用いることで、著作財産権の制約を解消し、更に、編曲、演奏、歌唱、録音することで、著作隣接権の制約を解消して、音楽データベースを構築する必要がある。ただし、著作者や権利者が特別に許可している楽曲では、その限りでない。また、いずれの場合も、研究目的での楽曲改変などが想定されることから、著作人格権 (同一性保持権など) を著作者に行使されない制作条件となっていることが望ましい。

1-7-2 音楽データベースの事例

研究目的での自由な利用のために構築された世界で最初の大規模な音楽データベースである「RWC (RealWorld Computing) 研究用音楽データベース」を紹介する。これは楽曲と楽器

音を共に含むものであるが、楽器音のみに関してはほかにも McGill University Master Samples, Iowa Musical Instrument Samples などの楽器音データベースが存在するので、それらを紹介する。また、クラシック音楽を中心に、楽譜をどう演奏するかという演奏表情に関するデータベースも構築されているので紹介する。

(1) RWC 研究用音楽データベース

「RWC 研究用音楽データベース」^{1,2,3)}は、「ポピュラー音楽データベース」(100曲)、「著作権切れ音楽データベース」(15曲)、「クラシック音楽データベース」(50曲)、「ジャズ音楽データベース」(50曲)、「音楽ジャンルデータベース」(100曲)、「楽器音データベース」(50楽器)の六つから構成される。全315曲に対し、音響信号、標準MIDIファイル、歌詞のテキストファイルが提供され、50楽器に対し、音域全体を半音間隔で収録した単独演奏音(楽器数約50種150個体で約29.1Gbytesの規模)を提供されている。

楽曲に関しては、理想的には、様々なジャンルにおける豊かなバリエーションをもつ楽曲を、高品質で可能な限り大量に収録することが望ましい。しかし、データベース制作上の資源の制約から限界がある。そこで、代表的な音楽ジャンルとして、ポピュラー音楽、クラシック音楽、ジャズ音楽の三つを取り上げ、「ポピュラー音楽データベース」(100曲)、「クラシック音楽データベース」(50曲)、「ジャズ音楽データベース」(50曲)が構築された。更に、有名楽曲集として、童謡を中心とした著作権の切れた楽曲を収録した「著作権切れ音楽データベース」(15曲)が構築され、音楽ジャンルの多様性を確保するために、少数ずつではあるが世界中の多様なジャンルの楽曲を収録した「音楽ジャンルデータベース」(100曲)が構築された。これら計315曲の楽曲は、実世界の音楽がもつ複雑さを反映し、可能な範囲でバリエーションが豊かとなるように制作された。例えば、曲調、テンポ、曲の長さ(曲長)、編成などが多様なだけでなく、作曲家、作詞家、編曲家、アーティスト(歌手)、楽器演奏者の人数が、制作資源の許す範囲で多くなっている。

一方、楽器音に関しては、弦楽器、管楽器、鍵盤楽器、打楽器、民族楽器などの多様な楽器について、その楽器で発音可能な音域全体を半音間隔で弾いた単独演奏音を収録した「楽器音データベース」(50楽器)が構築された。そして、奏法、強弱変化、楽器メーカー、奏者のバリエーションを豊かにして、可能な限り大量に収録されている。

研究者が研究目的で自由に使用できるようにするためには、楽曲に関する必要な著作権・著作隣接権が確保されている必要がある。そこで、著作権に関しては、新規に作曲、作詞、編曲した楽曲や、著作財産権の保護期間が終了した既存曲が用いられた。著作隣接権に関しては、すべての楽曲や楽器音が新規に演奏、歌唱、録音された。

RWC 研究用音楽データベースは、研究用途で利用可能となるよう広く配布しており、WWWサーチエンジンなどで、「RWC 研究用音楽データベース」を含むページを検索すると入手案内にアクセスできる。本データベースは既に国内外300機関以上に配布されて広く利用されており、音楽情報処理の研究分野の発展に寄与している。

(2) 楽器音データベース

楽器音を収録したデータベースは、上記の楽曲を収録したデータベースよりもずっと以前から構築されてきた。代表例として、「McGill University Master Samples」⁴⁾と「University of Iowa Musical Instrument Samples」⁵⁾の二つが挙げられる。

「McGill University Master Samples」⁴⁾は、1987年にFrank OpolkoとJoel Wapnickらによっ

て収録された楽器音のデータベースである。多様な楽器の単独音をステレオ収録した音響信号は、当初は11枚のCDとして配付されていたが、のちにすべての収録音が3枚のDVDとして配付されている。

一方、「University of Iowa Musical Instrument Samples」⁵⁾は、1997年にLawrence Frittsらによって収録された楽器音のデータベースである。多様な楽器の単独音をモノラル収録（ただしピアノはステレオ収録）しており、すべての収録音はWeb上からダウンロードして入手可能である。

(3) 演奏表情データベース

主にクラシック音楽では、作曲者が作曲した同一の楽譜が、様々な演奏者によって異なる演奏表情（テンポやタイミング、音量など）で演奏される。そこで、著作権の切れたピアノ楽曲を中心に構築された演奏表情データベースの代表例として、「CrestMusePEDB (CrestMuse Performance Expression Database)」^{6,7)}と「CHARM Mazurka Project」⁸⁾の二つが挙げられる。

「CrestMusePEDB」^{6,7)}は、CrestMuseプロジェクトによる音楽演奏表情データベースであり、市販されていて知名度の高いピアノの名演奏の音響信号を対象に、その演奏表情を手作業でラベル付けした結果が公開されている（音響信号は、同じ演奏が購入できることを前提に公開されていない）。具体的には、楽譜、演奏表情（機械的演奏からのずれ幅など）、音楽構造（階層的なフレーズ構造など）等のデータが、MusicXMLに準拠した形式で記述されている。更に、本データベース用に複数のピアノ奏者が新たに弾いた演奏音を録音した音響信号と、その演奏表情のデータも公開され、Web上からダウンロードして入手可能である。

「CHARM Mazurka Project」⁸⁾は、CHARMプロジェクトによる音楽演奏表情データベースであり、マズルカ (Mazurka) と呼ばれる形式の楽曲について、過去の演奏の音響信号を収集し、その演奏表情（主にビート位置など）を手作業や自動処理でラベルづけした結果が公開されている。拍のタイミング、テンポ、音量などについての詳細な分析結果も提供されている。

(4) 研究開発で利用されているその他の楽曲

音楽データベースというかたちはとっていないため、上記で紹介したデータベースのように研究目的で自由に利用可能なことが確保されているわけではないが、クリエイティブ・コモンズ・ライセンス⁹⁾の楽曲や、レコードレーベルMagnatune¹⁰⁾の楽曲などのようなWeb上でアクセス可能な楽曲は増え続けており、研究目的で利用されることがある。CGM（消費者生成メディア, Consumer Generated Media）あるいはUGC（ユーザ作成コンテンツ, User Generated Content）と呼ばれるエンドユーザによって創作された音楽も、利用されることがある。今後もそうしたコンテンツは増え続けることが期待できる。

1-7-3 音楽データベースに対するアノテーション

音楽データベースの内容に関するアノテーション（例えば、ビートやメロディ、コード、サビなど）は、機械学習、評価、比較、ベンチマークなどの目的で有用である。特に評価で用いるような正確なアノテーション（正解ラベル, Ground Truth）が求められる場合には、人手で付与することが多い。しかし、そうした作業は多大な労力がかかるため、アノテーション結果を研究コミュニティとして共有し、分野全体の発展に貢献する活動も始まっている。

例えば、上記のRWC研究用音楽データベースに対しては、同データベースを中心となっ

て構築した後藤真孝らによって、315 曲の全楽曲へ「AIST アノテーション」¹¹⁾ が付与され、データベース利用者に配付されている。四分音符レベルのビート位置と小節線からなるビート構造 (315 曲分)、10 ms ごとのメロディ音高 (F0) からなるメロディライン (115 曲分)、A メロ、B メロ、サビなどの楽曲構造 (300 曲分)、音響信号と同期した標準 MIDI ファイル (100 曲分) がダウンロード可能である。

音響信号自体を共有できない市販の楽曲に対しても、アノテーションだけを研究コミュニティで共有する活動が盛んである。例えば、アーティスト Beatles の全 12 枚のスタジオアルバム中の楽曲 179 曲に対するコード進行のアノテーション¹²⁾ が Christopher Harte らによって配付されている。これはその後、「OMRAS2 Metadata Project」¹³⁾ として拡大し、ビートやコード、調、楽曲構造が Beatles を含むほかの曲に対しても公開されている。また、ほかの研究機関でも様々な取り組みが行われている。

国際音楽情報検索学会の国際会議 ISMIR の場で、2005 年から始まった音楽情報検索関連技術のコミュニティでの評価の枠組み MIREX (Music Information Retrieval Evaluation eXchange)¹⁴⁾ やその前身の 2004 年の ADC (Audio Description Contest)¹⁵⁾ においても、様々なアノテーション (Ground Truth) が収集・利用されている。その一部は評価だけでなく、研究コミュニティで共有され、研究開発でも利用されている。

なお、アノテーションのためのツールの説明等を含む、より一般的なメタデータ/アノテーションに関する解説は、[1-8 メタデータ/アノテーション (梶)] 節に述べられている。

1-7-4 音楽データベースの今後

本節では、様々な音楽データベースの中から、研究開発目的で実際に多くの研究者によって利用されている楽曲、楽器音、演奏表情、アノテーションを中心に紹介した。音声や画像などのほかの分野でも、古くから共通データベースが構築されて分野の進展に貢献してきたように、研究分野が健全に発展していくうえで音楽データベースは重要である。今後も、世界中で音楽情報処理のための多様なデータベースが構築され、それに基づいて分野全体が大きく進展していくことが期待される。

■参考文献

- 1) 後藤真孝, 橋口博樹, 西村拓一, 岡隆一, “RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース,” 情報処理学会論文誌, vol.45, no.3, pp.728-738, 2004.
- 2) M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, “RWC Music Database: Popular, Classical, and Jazz Music Databases,” Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR 2002), pp.287-288, 2002.
- 3) M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, “RWC Music Database: Music Genre Database and Musical Instrument Sound Database,” Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003), pp.229-230, 2003.
- 4) F. Opolko and J. Wapnick, McGill University Master Samples (CDs), 1987.
- 5) L. Fritts, “University of Iowa Musical Instrument Samples,” <http://theremin.music.uiowa.edu/MIS.html>
- 6) 橋田光代, 松井淑恵, 北原鉄朗, 片寄晴弘, “ピアノ名演奏の演奏表現情報と音楽構造情報を対象とした音楽演奏表情データベース CrestMusePEDB の構築,” 情報処理学会論文誌, vol.50, no.3, pp.1090-1099, 2009.
- 7) M. Hashida, T. Matsui, and H. Katayose, “A New Music Database Describing Deviation Information of

- Performance Expressions,” Proceedings of the 9th International Conference on Music Information Retrieval (ISMIR 2008), pp.489-494, 2008.
- 8) CHARM Mazurka Project, <http://www.mazurka.org.uk/>
 - 9) Creative Commons, <http://creativecommons.org/>
 - 10) Magnatune, <http://www.magnatune.com/>
 - 11) Masataka Goto, “AIST Annotation for the RWC Music Database,” Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR 2006), pp.359-360, 2006.
 - 12) Christopher Harte, Mark Sandler, Samer Abdallah, and Emilia Gomez, “Symbolic representation of musical chords: A proposed syntax for text annotations,” Proceedings of the 6th International Conference on Music Information Retrieval (ISMIR 2005), pp.66-71, 2005.
 - 13) M. Mauch, C. Cannam, M. Davies, S. Dixon, C. Harte, S. Kolozali, D. Tidhar, and M. Sandler. “OMRAS2 Metadata Project 2009,” Late-breaking poster of the 10th International Conference on Music Information Retrieval (ISMIR 2009), 2009.
 - 14) J.S. Downie, “The music information retrieval evaluation exchange(2005-2007): A window into music information retrieval research,” Acoustical Science and Technology, vol.29, no.4, pp.247-255, 2008.
 - 15) P. Cano, E. Gomez, F. Gouyon, P. Herrera, M. Koppenberger, B. Ong, X. Serra, S. Streich and N. Wack, “ISMIR 2004 Audio Description Contest,” MTG Technical Report MTG-TR-2006-02, 2004.

■2群-9編-1章

1-8 メタデータ・アノテーション

(執筆者：梶 克彦) [2011年6月 受領]

コンテンツの意味内容(コンテンツ自体の構造, 制作者が込めた意図など)や周辺情報(制作者情報, 関連コンテンツなど)に基づいて検索, 推薦, 変換などのタスクを行うためには, それらの情報を計算機が処理可能な形式で記述し, 管理する仕組みが必要である. そこで, コンテンツの様々な情報を注釈として関連付けられるアノテーション(メタデータとも呼ばれる)に関する研究やサービスが盛んになってきている.

1-8-1 主なアノテーション記述形式

現在, 音楽におけるアノテーションとして最も一般に普及しているのが ID3 タグ¹⁾である. ID3 タグは, アーティスト, 曲名, ジャンルなど, 楽曲の基本情報をオーディオファイルの形式である MP3 に埋め込むための規格である. 楽曲の基本情報を管理する CDDB²⁾が ID3 タグの普及を促進した. 多くの音楽プレイヤーは, CDDB から楽曲情報を自動的にダウンロードし MP3 に ID3 タグを埋め込む機能をもっている.

埋め込み式のアノテーションは, コンテンツと同時に流通するため利便性が高いという利点があるが, アノテーション管理を一元化することができず記述形式の柔軟性に問題がある. そこで, コンテンツとは分離して記述可能なアノテーション形式が提案されている.

コンテンツと分離したアノテーションの記述形式には, 記述の柔軟性から XML が用いられることが多い. MusicXML³⁾, WEDELMUSIC XML Format⁴⁾は楽曲の楽譜相当の情報を記述するための形式である. また MPEG-7⁵⁾は音声や動画などのマルチメディアコンテンツに対するアノテーション記述形式である. オーディオ信号の特徴などの低レベルな情報から, 楽器の音色や音声認識の結果などの高レベルな情報まで記述可能である. CrestMuseXML⁶⁾は音楽情報処理研究のための共通データフォーマットである. 音楽情報処理によって得られる様々な結果を, リンクされた複数の XML ドキュメントとして表すことができる.

共通した形式のアノテーションを Web 上で公開することで, 多くのユーザがアノテーションの検索や応用を行うことができるようになる. そこで, コンテンツに共通する基本情報(タイトル, 制作者, 提供者, 言語など)の語彙として Dublin Core⁷⁾が提唱された. また RDF (Resource Description Framework) によって, Web 上のリソースに対するアノテーションの記述形式が共通化された. RDF はリソース間の関係を有向グラフとして記述する. これによりコンテンツ間の複雑な関係を表現することが可能になる. MusicBrainz⁹⁾はアーティストや作品の基本情報やその関係などをデータベース化するプロジェクトである. Dublin Core や RDF を採用し, 更に音楽特有の語彙(アルバム, トラックの長さ, 楽曲指紋など)の定義を行っているため, 拡張性, 汎用性が高い形式であるといえる.

1-8-2 アノテーションシステム

Marsyas¹⁰⁾は楽曲オーディオファイルの音響解析結果をアノテーションとして自動解析するシステムである. パワースペクトル, MFCC, 音量, 音高, 周期性など, 多くの音楽研究

に有効な解析結果を提供する。本システムはスタンドアロンで動作し、コマンドラインかGUIにより操作を行う。

音響信号の可視化により、アノテーションを容易にしたのが Clam Music Annotator¹⁾ や Sonic Visualiser¹²⁾ である。Clam Music Annotator (図 1・4) は、時間範囲に対する音高や和音などの情報を付与または編集することができる。Sonic Visualiser は波形やスペクトルなどの音響信号の解析結果をアノテーションとしてインポート、エクスポート可能なのが特徴である。

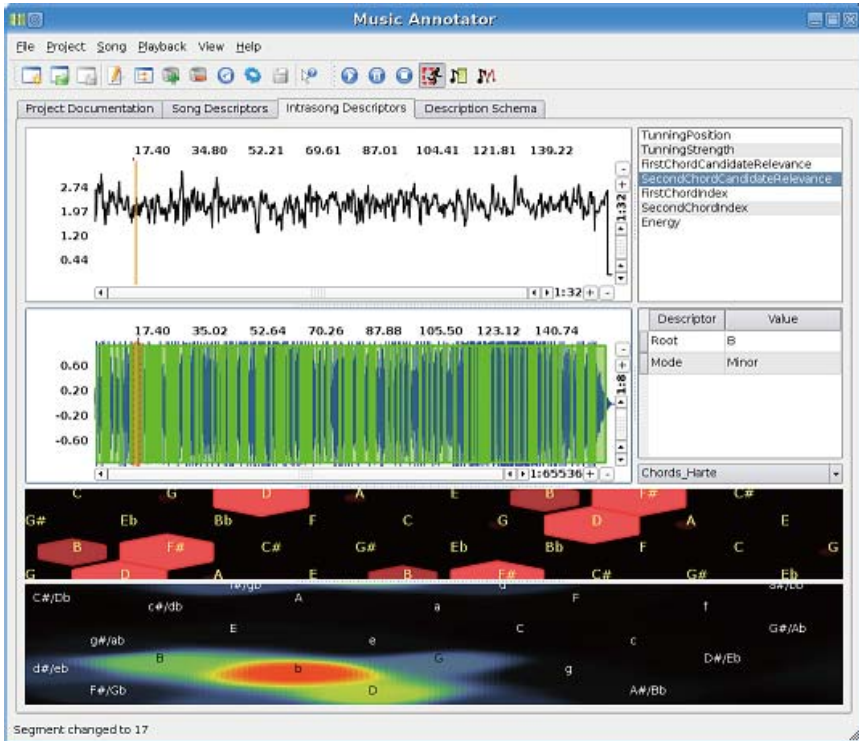


図 1・4 Clam Music Annotator

大量のアノテーションを獲得するためには、Web 上のユーザから収集することが有効である。大衆的な音楽サービスを提供するためには、膨大な数の楽曲に対するアノテーションを網羅的に収集する必要がある。そこで CDDDB は、未登録の楽曲基本情報をインターネット上のユーザから音楽プレイヤーを通して収集している。

一般ユーザがアノテーションにかかわることによるノイズの問題がある。正確なアノテーションを収集するためには、ユーザに何らかのインセンティブを与える必要がある。そこでゲームによるアノテーション手法が注目を集めている。Google Image Labeler²⁰⁾ は The ESP

Game²¹⁾を基にしたゲームである。2人の同じ画像を見ている人が同じタグを付与したらスコアが加算される。このようなゲームの手法は音楽アノテーションにも取り入れられている^{22,23)}。

一般にサービスプロバイダがトップダウンに提供する検索項目と、ユーザが検索したい項目は必ずしも一致しない。そこでフォークソノミーというユーザが付与するタグによる分類法が注目を集めている。フォークソノミーに基づくサービスでは、ユーザがコンテンツに対して自由に複数のタグを付与でき、コンテンツはユーザに付与されたタグ集合により特徴づけられる。ユーザはタグクラウド(コンテンツのタグ集合を一挙に表示する可視化手法)を見ることでそのコンテンツの内容を推測したり、タグをたどって類似コンテンツを検索することが可能である。既にAnnotea¹⁴⁾, del.icio.us¹⁵⁾, flickr¹⁶⁾, youtube¹⁷⁾など多くのシステムで取り入れられている。音楽サービスとしてはLast.fm¹⁸⁾が代表的であり、楽曲やアルバム、アーティストに対して、ジャンルや印象、聴きたい状況など自由にタグを付与することができる。

フォークソノミーには、様々な属性の語がタグという種類のアノテーションとして付与されることによる分類の粗さの問題がある。そこでフォークソノミーの柔軟性をもち、かつアノテーションの詳細な分類が可能である音楽アノテーションシステム¹⁹⁾が提案された。Web上のユーザから図1・5のように楽譜中に現れる音符や歌詞などの要素集合に対するアノテーションを収集する。収集するアノテーションの種類(解説, 印象, コードなど)は、ユーザにより自由に設定可能であり、ほかのユーザが作成したアノテーションの種類を用いることもできる。

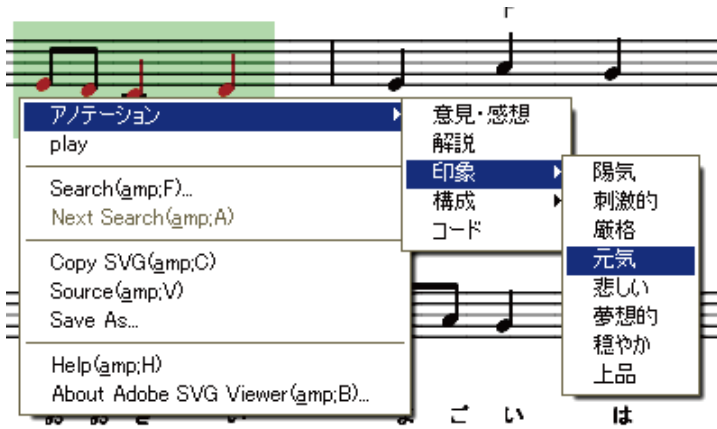


図1・5 音楽アノテーションシステム

1-8-3 応用

楽曲鑑賞にアノテーションを利用する例として、MoodLogic²⁴⁾とMusicream²⁵⁾を挙げる。MoodLogicはWeb上のユーザコミュニティと専門家によってテンポや印象情報を収集し、「アップテンポなロック」や「ロマンチックなブルース」など、**印象情報に基づく検索**やプレイ

リスト生成が可能である。図 1・6 に示す Musicream は「流し楽曲」「類似くつつき」「メタプレイリスト」「タイムマシン」といった機能による新しい音楽の聴き方を提供するシステムであり、Marsyasの解析結果に基づいて実装されている。

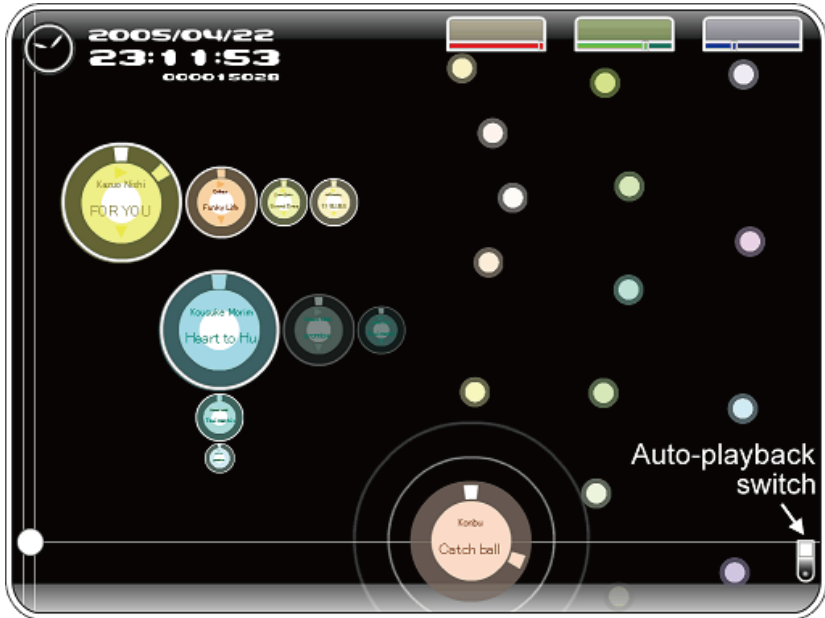


図 1・6 Musicream

次に楽曲制作にアノテーションを利用する例として CUIDADO (Content-based Unified Interfaces and Descriptors for Audio/music Databases available Online)²⁶⁾ を挙げる。CUIDADO は楽曲ブラウジング・検索・オーサリングの支援を目指すプロジェクトである。音楽制作者のための検索・編集・処理ツールである Sound Palette が提供されている。ユーザは楽曲ファイルをインポートするだけで、自動でセグメント化された楽曲断片が利用可能になる。またテンポの異なる二つの楽曲をミックスさせる際に、自動的にテンポを同期させるなど、音楽のセマンティクスに基づいた編集を支援する。

Web 上のユーザとのコミュニケーションの場を提供するシステムとして、Pandora, Last.fm, を挙げる。音楽のコミュニケーションサービスとして既に多くのユーザを集めているのが、前述の Pandora や Last.fm といった SNS (ソーシャルネットワーキングサービス) である。SNS とはインターネットを介して人と人のコミュニケーションをサポートするサービスである。これらのシステムでは、楽曲の鑑賞履歴や楽曲の好き嫌いのフィードバック情報を用いてユーザプロフィールを作成し、協調フィルタリングによって類似する楽曲を好むほかのユーザを発見し紹介する。

■参考文献

- 1) ID3v2, <http://www.id3.org/>
- 2) Gracenote, CDDB, <http://www.gracenote.com>
- 3) M. Good, "MusicXML: An Internet-Friendly Format for Sheet Music," Proceedings of XML 2001 Conference, 2001.
- 4) P. Bellini and P. Nesi, "WEDELMUSIC Format: An XML Music Notation Format for Emerging Applications," Proceedings of WEDELMUSIC, pp.79-86, 2001.
- 5) B.S. Manjunath, P. Salembier, and T. Sikora, "Introduction of MPEG-7," John Wiley&SonsLtd., 2002.
- 6) 北原鉄朗, 橋田光代, 片寄晴弘, "音楽情報科学研究のための共通データフォーマットの確立を目指して," 情報処理学会研究報告, 2007-MUS-71, pp.149-154, 2007.
- 7) Dublin Core Metadata Initiative, <http://dublincore.org/>
- 8) Resource Description Framework(RDF), <http://www.w3.org/RDF/>
- 9) A. Swartz, "MusicBrainz: A Semantic Web Service, Intelligent Systems," IEEE, vol.17, Issue 1, pp.76-77, 2002.
- 10) G. Tzanetakis and P. Cook, "MARSYAS: A Framework for Audio Analysis Organized Sound," Cambridge University Press, vol.4, no.3, 2000.
- 11) X. Amatriain, J. Massager, D. Garcia and I. Mosquera, "The CLAM Annotator: A Cross-platform Audio Descriptors Editing Tool," Proceedings of ISMIR, 2005.
- 12) C. Cannam, C. Landone, M. Sandler and J.P. Bello, "The Sonic Visualizer: A Visualization Platform for Semantic Descriptors from Musical Signals."
- 13) Pandora Internet Radio, <http://www.pandora.com/>
- 14) J. Kahan, M. Koivunen, E. Prud'Hommeaux and R. Swick, "Annotea: An Open RDF Infrastructure for Shared Web Annotations," Proceedings of WWW, 2001.
- 15) del.icio.us, <http://del.icio.us/>
- 16) Flickr, <http://www.flickr.com/>
- 17) YouTube, <http://jp.youtube.com/>
- 18) Last.fm, <http://www.lastfm.jp/>
- 19) 梶 克彦, 長尾 確, "楽曲に対する多様な解釈を扱う音楽アプリケーションシステム," 情報処理学会論文誌, vol.48, no.1, pp.258-273, 2007.
- 20) Google Image Labeler, <http://images.google.com/imagelabeler/>
- 21) L.v. Ahn and L. Dabbish, "Labeling Images with a Computer Game," Proceedings of CHI, pp.319-326, 2004.
- 22) E.L.M. Law, L.v. Ahn, R.B. Dannenberg, M. Crawford, "TagATune: A Game for Music and Sound Annotation," Proceedings of ISMIR, 2007.
- 23) M.I. Mandel, D.P.W. Ellis, "A Web-Based Game for Collecting Music Metadata," Proceedings of ISMIR, 2007.
- 24) MoodLogic, <http://www.moodlogic.com/>
- 25) M. Goto and T. Goto, "Musicream: New Music Playback Interface for Streaming, Sticking, Sorting, and Recalling Musical Pieces," Proceedings of ISMIR, 2005.
- 26) H. Vinet, P. Herrera and F.F. Pachet, "The CUIDADO Project," Proceedings of ISMIR, 2002.

■2 群-9 編-1 章

1-9 音楽配信

(執筆者：鈴木健嗣) [2011年6月 受領]

音楽配信とは、一般に携帯電話などのモバイル端末、及びインターネットを通じて楽曲を配信することを指す。携帯電話などのモバイル端末、及びインターネットを通じた国内の有料音楽配信売上高は754億円(2007年¹⁾)であり、急速に市場を伸ばしている音楽産業の一分野である。近年、音楽のみならず、写真、映像といったコンテンツのデジタル化が進み、ネットワークに係る通信技術の進歩、高速通信・高スループット環境の普及を実現するインフラ整備に伴い、高音質高画質のコンテンツがネットワークを介して配信されている。

また、インターネットラジオやポッドキャストに代表される新しい放送メディア方式の普及や、計算機やモバイル端末におけるメディア再生機器の小型化と性能向上とともに、時間と場所を問わずに自由に多様な楽曲を取得し、楽しむことが一般的になってきた。このように、従来のレコードやCDといった媒体を利用したパッケージビジネスと比較しても、ネットワークを通じた音楽・映像配信はコンテンツ流通ビジネス²⁾の中核になり、現在広く普及している。その一方、著作権保護、及び有料配信に係る課金システムはいまだ成熟しておらず、特に違法コピーや不正流通を防止するための技術や関連の法整備が急務である。

そこで本節では、音楽配信に係る技術を概説するとともに、デジタル著作権管理、及び音楽配信市場の推移について述べる。

1-9-1 音楽配信と通信技術

音楽の配信形式は、ストリーミング配信、ダウンロード方式の二つに大別される。

使用者の端末や計算機にファイルを保存するダウンロード形式に対し、ストリーミング(streaming)とは、音楽・音声・動画などのコンテンツをストリーミングサーバから転送・再生する方式³⁾であり、逐次の再生している時刻の直近のみバッファされるコンテンツを視聴するため、通常ファイルを視聴者側の端末に保存しない。これらのデータストリームをリアルタイムに配信するため、データ転送プロトコルRTP(Real-time Transport Protocol)が用いられる。RTPはUDPの上位プロトコルであり、データ喪失や、配送遅延がある。このRTPを利用して、同時にフロー制御のためにデータ品質や送信者・受信者情報などの制御情報を通信するRTCPプロトコルや、ビデオ会議を含むマルチメディア通信用のH.232形式などが音楽映像配信に広く用いられている⁴⁾。なお、RTPプロトコルは、IETF(The Internet Engineering Task Force)やITU(International Telecommunication Union)にて標準化されている。このようにストリーミング方式は、音楽コンテンツといった比較的大きなサイズのファイル再生に有効であるだけでなく、無断で複製することが困難なため、課金システムの観点からも有効である。更に、端末側にファイル保存のための大きな容量を必要としないため、携帯電話やPDAにてパケットストリーミングで音楽を配信するサービスが行われている。また、演奏会やコンサート会場からのライブ配信などコンテンツの制作者から視聴者へ直接音楽を配信することが可能である。しかしながら、再生品質や配送遅延が通信路の状態に依存するため、安定音楽配信についてはいまだ課題がある。

また近年、プログレッシブダウンロードと呼ばれる配信方法も広く普及している⁵⁾。これは技術的にはダウンロード方式の一部と位置づけられるが、flv形式に代表されるように、ダウンロード中においても再生を可能にした形式であり、再生開始に必要な制御情報がファイル先頭部に記述されていることが特徴であり、データファイルのダウンロード開始直後から音楽を楽しむことが可能である。またこの方式は、HTTPを通じて配信を行い、接続速度の異なる視聴者にも同品質のコンテンツを提供することが可能である。

インターネットプロトコル(HTTP)を通じ、ストリーミング方式にて音声や音楽などの番組配信を行う媒体として、インターネットラジオがある。特に国外で顕著であるが、既存の放送局が従来の電波放送と同じ内容の音声・音楽を放送しているものも多い。また、Windows Media PlayerやiTunesといった再生ソフトウェアでは、標準でインターネットラジオ利用機能が搭載されており、これもネットラジオ普及に重要な役割を果たしている。また、ダウンロード方式では、ポッドキャスト方式による放送が目目され、インターネット放送を専門とする企業や団体や個人による放送のみならず、既存の大手放送局も参入し始めている。

一方、携帯電話を代表とする電波を用いた無線通信機器の普及と機能向上を背景として、移動体端末への音楽配信も大きな産業となっている。従来の無線通信機器では、インターネット上の情報にアクセスするためのプロトコルであるWAP(Wireless Application Protocol)が独自の規格に基づいているために、TCP/IPやHTMLとの親和性が低かったが、2001年に新たにWAP2.0が規格化され、標準的なインターネット規格であるとの親和性が大きく高まった⁶⁾。これにあたり、音声、画像、テキスト情報を統合した情報の送受信が可能なMMS(Multimedia Messaging Service)機能が導入され、無線通信機器も本格的に音楽配信の対象となった⁷⁾。特に日本では、2008年現在携帯電話を使った音楽配信が広く普及しており、計算機でのダウンロード利用より大きな産業となっているのが特徴である。

特に、楽曲全体を配信するのではなく、元の楽曲の音源から一部を抽出し、15~30秒程度とした音楽ファイルを携帯電話にダウンロードして、この楽曲を着信音の代わりに利用するサービスが提供されている。これらは、携帯電話端末によるコンテンツ検索、購入、ダウンロードから課金システムまでが整備されており、ユーザビリティの高さから広く普及している。

1-9-2 符号化方式・圧縮形式とデジタル著作権管理

非圧縮音声・音楽ファイルの基本フォーマットは、音声・音楽などのアナログ信号を特定のビット深度にて標本化し、パルス列に変換するパルス符号変調(PCM:pulse code modulation)に基づいている。また、符号化方式(コーデック)とは、この音声情報のデジタルデータへの変換方式であり、通常各コーデックにそれぞれのファイルフォーマットが対応する。このような音声・音楽ファイルのフォーマットは、非圧縮音声フォーマット、可逆圧縮フォーマット、非可逆圧縮フォーマットの三つに分類されるが、音楽配信では効率良い圧縮形式が採用される場合が多く、特にMP3(MPEG Audio Layer-3)が標準的に用いられる。これは、MDCT(修正離散コサイン変換)による音声圧縮アルゴリズムを用いており、高い圧縮率を有するのが特徴である⁸⁾。

音楽配信に使われる主な音声フォーマットは、AAC(アップル社)、WMA(マイクロソフト社)やATRAC3(株式会社ソニー)である。これらはすべて非可逆圧縮の音声フォーマット

トであるため、音質の劣化は避けられない。また、多くの音声フォーマットには、デジタル著作権管理機能や違法コピー防止技術が含まれており、コンテンツ配信企業や利用するハードウェア、及び課金システムによって異なるフォーマットとなっているのが一般的である。このため、音声フォーマットの評価においては、符号化方式・圧縮方式の効率や性能のみの議論に留まらないことが特徴である。

このような電子機器上のコンテンツに対し、第三者による無制限な複製や再利用を防止することを目的として、デジタル著作権管理 (DRM: Digital Rights Management) の技術が注目されている⁹⁾。従来より、著作権者がその著作物の無断複製を防止するために、メディアや録音機器にコピーガード技術が適用されてきた。例えば、録音機器に対する SCMS (Serial Copy Management System) は、デジタル録音機器のコピー防止技術であり、デジタル接続によるコンテンツの複製が一世代までに制限されるものである。これは、機器がデジタル信号に含まれるコピー情報を検出し、録音を行わない仕組みである。しかしながら、アナログ接続による複製には効果がないため、無制限で複製することが可能である。これに対し DRM は、ハードウェア的な複製防止でなく、暗号化されたデジタルコンテンツとそれを再生するソフトウェアのみからなるものであり、ファイルの複製は可能であってもそれを再生・閲覧を制限することが可能である。DRM を使用した音楽配信では、購入する楽曲ごとに、移動端末や CD への転送・書込み回数制限を設定することが可能である。ただし、複製以外の利用として、著作権法の範囲内での抜粋、及び他人への譲渡も制限されてしまうなど、個人の私的利用まで制限されるといった問題がある。また、DRM の暗号化や再生ソフトウェアの仕様は公開されておらず、将来に渡って視聴者の権利が担保されないといった問題もある。

このような著作権保護技術の統一方式の確立を目指し、SDMI (Secure Digital Music Initiative) といった国際的な業界団体¹⁰⁾があり、そこで音楽配信の標準セキュアフォーマットを定める技術的仕様を提供している。

1-9-3 音楽配信市場の推移

日本では、1999年より電子透かし技術などのデジタル著作権管理を考慮した音楽配信サービスが行われていた。しかしながら提供される楽曲の少なさや、既存の販売媒体と比べて優位ある価格でなかったことより、普及には至らなかった。当時、まだ音楽配信市場が拡大する以前は、ユーザによる無料の音楽ファイルの交換が大きな社会的問題となっていた。ここでは、特定のサーバを介さず、ユーザどうしの直接接続によるファイル交換を実現するシステムが開発され、違法性を伴うファイル交換が行われていた。ところが2001年には、米サンフランシスコ連邦高等裁判所により、インターネット上で無料音楽交換サービスを展開していた企業に対し、レコード会社が保有する著作権を著しく侵害しているとの判断より、事業停止を勧告した。ここでは、レコード業界における CD などの売上減少要因と認定され、無断複製と大量送信行為の補助として違法行為との判断となった。日本でも、同様のファイル交換サービスを行っていた企業に対し、日本音楽著作権協会 (JASRAC) と日本レコード協会 (RIAJ) 加盟会社が共同で、同社のサービス停止を求めて、サービス停止を求める仮処分申請を行った¹¹⁾。その結果、当該企業が著作権侵害の主体と認定され、やはりサービス停止の処分となっている。

その後2003年、米アップル社がiTunes Music Storeを開設し、携帯型デジタルオーディオプレイヤーの急速な普及、配信する楽曲数と価格の優位性を背景として、それまで拡大していなかった有料音楽配信の市場にて大きな収益を上げた。その後、インターネット企業やレコード業界からも多くの企業が有料の音楽配信市場に参入し、独自の音楽配信事業を行っている。日本レコード協会の報告では、既に日本では、2005年携帯端末や計算機による音楽配信の売上は、シングルCDの売上を上回っており¹²⁾、世界的にもダウンロード・ストリーミングによる音楽配信に係る売上は、7%以上¹³⁾となっている。このように、音楽配信市場の市場規模も大きく拡大しており、今後も引き続き成長の見通し¹⁴⁾である。

このようにネットワークを通じた音楽配信を利用すると、デジタル化により音楽検索が容易になったため、特定の楽曲のみが欲しい場合や、廃盤となったCDの楽曲など入手困難なものを入手できる可能性があるなど、視聴者にとって大きな利点が多い。また、これまで既存の企業や特定のアーティストのみが音楽配信を行っていたが、インターネットラジオやポッドキャスト方式等による新しい音楽配信システムを利用することで、個人が手軽に放送局を開局できるだけでなく、世界中に楽曲を配信することが可能である。また特に国外では多くの放送局がインターネットラジオを通じて音楽を配信しているが、いまだ課金システムが未成熟のため産業としては広まっていない。しかしながら、このような国際的な情報発信に対する著作権問題を改善するために、利用者の利便性を担保しつつ、違法な複製や利用を制限するような新たな法整備が急務である。以上のとおり、今後も音楽配信市場は着実に成長すると予想されるが、著作権管理とプライバシーの保護という課題に取り組んでいかなければならない。

■参考文献

- 1) 一般社団法人 日本レコード協会 有料音楽配信売上実績,
<http://www.riaj.or.jp/data/download/2008.html>
- 2) 関亜紀子, 亀山 渉, “コンテンツ流通ビジネスのモデル化と評価に関する検討,” IPSJ SIG Notes 2004, vol.27, pp.23-30, 2004.
- 3) “通信・ネットワーク事典,” 日経コミュニケーション版, 2003.
- 4) “最新情報通信用語集,” 電気通信協会, 2000.
- 5) Stefan Richter, Jan Ozer, “Hands-On Guide to Flash Video: Web Video and Flash Media Server,” Focal Press, 2007.
- 6) WAP Forum, <http://www.wapforum.org/>
- 7) 3rd Generation Partnership Project (3GPP), <http://www.3gpp.org/>
- 8) MPEG (Moving Picture Experts Group), ISO/IEC JTC 1/SC 29/WG 11, <http://jtc1.org/>
- 9) 平成17年特許出願技術動向調査報告書 デジタル (DRM) 著作権管理, 特許庁, 2006.
- 10) SDMI (Secure Digital Music Initiative), <http://www.sdmi.org/>
- 11) 黒田特許事務所, “図解でわかるデジタルコンテンツと知的財産権,” 2004.
- 12) “日本のレコード産業2005年度版,” 日本レコード協会, 2006.
- 13) Digital Music Report 2006, IFPI (International Federation of the Photographic Industry), 2006.
- 14) “デジタルコンテンツ白書2008,” 財団法人デジタルコンテンツ協会, 2008.

■2群-9編-1章

1-10 CGM (Consumer Generated Media)

(執筆者：平井重行) [2011年6月 受領]

ユーザもしくは消費者が生み出すデジタル情報をもとに新たなコンテンツやその価値を生み出すことを指す Web2.0 関連用語である。以前から、Web 上の掲示板やブログ、SNS などにおいて書き込まれたコメントやレビューなど、テキスト情報主体の「集合知」という概念で知られており、大手のコンテンツ販売サイトや SNS などでは CGM のマーケティングやビジネスへの活用が行われている。最近では、ポッドキャストや動画共有サイトなどの登場で、画像やムービー、音楽などのコンテンツを一般ユーザがアップロードする CGM も増えている。ここにおいては、プロの制作物ほどクオリティが高くない素人のものであっても、多くの人が興味をもって話題となり、ビジネスへと発展する場合も出てきている。この例のように、CGM はこれまで消費者であったユーザがコンテンツ提供者となることも意味している。

以下に、音楽音響メディアに関する CGM についていくつかの側面からまとめる。

1-10-1 コンテンツ販売・マーケティング活用

音楽CD/DVDや楽譜などを販売するWebサイト (Amazon¹⁾ やHMV²⁾ など) や、楽曲のダウンロード販売を行うサイト (iTunes Music Store³⁾ やmora win⁴⁾ など) においては、売れ筋アルバムや楽曲などのランキングを行うほか、購入者が書き込むコンテンツ・レビューを提示することによりCGMを実現している。別ユーザがコンテンツ購入時にこれら情報を参考にでき、過去の購入者が同時購入したコンテンツを紹介、推薦するなどCGMがマーケティングに大きく寄与する。基本的にはAmazonなどの大手コンテンツ販売サイトで多くの情報が集まるのでCGMとしてのマーケティングの効果が現れやすいが、インディーズや自作曲を扱うRecommuni⁵⁾のようにSNSと組み合わせることで口コミ的に推薦するかたちでCGMを活用することもある。

1-10-2 鑑賞楽曲リストの公開と鑑賞ランキング

Last.fm⁶⁾ や MyStrands⁷⁾、mixi music⁸⁾、プレイログ⁹⁾、音ログ¹⁰⁾ などの SNS では、iTunes や Windows Media Player、WinAmp、iPod などの音楽プレイヤーで、実際に再生された楽曲のメタデータを収集する機能をもっている。収集されたデータは、各ユーザの鑑賞楽曲リストとして他ユーザが閲覧できるほか、楽曲ごとやアーティストごと、ユーザグループごとで再生回数を集計してランキングする機能なども用意されている。これによって、各アーティストの人気楽曲の傾向や、アーティスト自身の人気、各ユーザの音楽嗜好が分かるなど、2 次的な情報を得ることができる。これらの情報をもとに楽曲やアルバムの推薦を行い、販売するサービスは一般的である。

1-10-3 自作コンテンツやプレイリストの公開環境

米国最大の SNS の一つ MySpace¹¹⁾ では、アーティストという分類のユーザであれば楽曲

の公開ができ、メジャーなアーティストからインディーズのほか、趣味で音楽制作している人まで楽曲を公開している。また、YouTube やニコニコ動画などの動画共有サイトにおいても、音楽に関する映像コンテンツが多数アップロードされ、自作コンテンツ公開の場として人気を博している。これら SNS や動画共有サイトにおいては、公開されたコンテンツにユーザが検索用タグを付けたり（フォークソノミー）、コメントやレビューを書き込むことができ、CGM プラットフォームとして活用されている。また、音楽系 SNS の Last.fm や Yahoo! Music Jukebox などのサービスでは、自分の好みの楽曲をまとめたプレイリストを作成してタグを付けて公開し、それらにコメントを付けることができる。SNS のこれらの機能は、場合によって楽曲やアルバムなどのコンテンツ販売サイトと連携してマーケティングに寄与することもある。このように、自らが楽曲やプレイリストを作成・公開し、SNS の機能と合わせることで CGM としての活用が行われている。

一方で、ポッドキャスト（ボイスブログやねとらじ、デジオ¹²⁾などの名称のものも含む）においては、自作楽曲を公開することも可能だが、それ以外のコンテンツも多く提供されている。例えば DJ 気分で制作したラジオ番組、様々な分野の講習内容を録音したコンテンツなどである。録音したファイルを MP3 としてアップロードするだけで、手軽にコンテンツを公開できる点が CGM としての素地であるといえる。ポッドキャストの ASP では、タグ付けが自由にできるほか、併設するブログや掲示板があるとそこから話題になって人気が出るコンテンツも多々ある。

1-10-4 コンテンツ所有リストの公開

CGM の対象となる様々なコンテンツ（情報や物品含む）の中で、音楽コンテンツは楽曲データや音楽 CD/DVD というかたちで収集、所蔵することが比較的多いものである。それゆえ、Delicious Monster¹³⁾ やブクログのように、個人所有の楽曲やアルバムのリストを作成して Web 上で公開するサービスがある。楽曲やアルバムなど所有コンテンツの数が膨大なユーザには、それらのコンテンツの所有リストを作成し、検索できる管理システムがあれば有益である。また、多数のコンテンツをもっているユーザの所有リストが公開されていれば、他のユーザが楽曲やアルバムを購入する際の参考になる。このことは、コンテンツ販売サイトにおける同時購入履歴や閲覧履歴に基づくコンテンツ推薦とは違う目的の情報として活用できる。

1-10-5 コンテンツ制作環境の整備

プロが使う制作環境でなくとも、また音楽知識や経験があまりなくとも、それなりのクオリティの作編曲が可能なソフトウェアが登場してきたことは、これまで音楽の鑑賞者であったユーザをコンテンツ制作者へと変化させる大きな要因として考えられる。その典型的なソフトウェアの例としては Apple 社の GarageBand が挙げられる。数多くのリフやフレーズ、リズムパターンのデータ素片をドラッグアンドドロップ操作によって配置するだけで曲が作成可能であり、これが Mac には標準で搭載されていることから誰でも容易に音楽コンテンツの制作ができる。

また、最近ではクリプトンフューチャー社の歌声合成ソフトウェア「初音ミク」及び「鏡音リン・レン」の登場も趣味の音楽制作に大きく貢献していると考えられる。これらは、ヤ

マハ社の歌声合成エンジン VOCALOID2 をベースに作られた製品であるが、それらの独特な声質と共に様々な曲を自由に歌わせることができる点で人気を呼び、趣味で作成したコンテンツをニコニコ動画や YouTube で公開することが一躍ブームとなった。

1-10-6 CGM としてのメディアアート

インスタレーションなどのアート作品によっては、鑑賞者が作品鑑賞する行為自体で音楽を生成するものもあり、その音楽を聞いたほかの鑑賞者がまた別の音楽を生成してゆくこととなる。このようなものも CGM の一つの形といえる。また、安齋らによるカンプリアン・ゲームを音に適用した作品「音の樹¹⁴⁾」では、インターネット上に携帯電話などで録った画像や音を投稿し、その音を聴いたユーザがそれに触発されて別の音を投稿することを多数ユーザによって繰り返し行われる。その結果、音によって触発されたアイデアやインスピレーションの連鎖で、成長する音の作品として成立している。これも携帯電話やインターネットを活用した CGM の一つと考えることができる。

このように、音や音楽に限った話ではないが、ユーザ参加型のメディアアート作品は CGM としての側面をもっているといえる。

■参考サイト

- 1) <http://www.amazon.co.jp/>
- 2) <http://www.hmv.co.jp/>
- 3) <http://www.apple.com/itunes/store/>
- 4) <http://morawin.jp/>
- 5) <http://recommuni.jp/>
- 6) <http://www.last.fm/>
- 7) <http://www.mystrands.com/>
- 8) <http://mixi.jp/>
- 9) <http://playlog.jp/>
- 10) <http://otolog.jp/>
- 11) <http://www.myspace.com/>
- 12) <http://ladio.net/>
- 13) <http://www.delicious-monster.com/>
- 14) <http://anzlab.com/st06a/>

■2 群-9 編-1 章

1-11 音楽生成言語・環境

(執筆者：平井重行) [2011年6月 受領]

アルゴリズムコンポジションや、リアルタイムパフォーマンスのための音楽・音響生成など、コンピュータ音楽の研究や作品制作に利用される音楽生成言語・環境には様々なものが存在する。それら是对応する言語の種類や実行プラットフォーム、専用ハードウェアの有無、用途、手軽さ、複雑さ、など様々な観点で見るとかなり多岐に渡る。言語の種類や記述方式の観点でそれらを分類すると次表のようになる。

表 1・1 音楽生成言語・環境の分類

| | |
|---------------|--|
| 手続き型言語系 | Csound, RTcmix, ChucK |
| オブジェクト指向型言語系 | SuperCollider, JSyn, jMusic, ChucK |
| 関数型言語系 | CLM, Haskore |
| ビジュアルプログラミング系 | Max/MSP, PureData, jMax, OpenMusic, Kyma X |
| データ記述系 | MUSIC V, Csound, MusicXML, SAOL, etc. |
| その他 | M, UPIC, etc. |

これらの言語や環境のうち、インタラクティブな音楽作品に対してはビジュアルプログラミング環境（特に Max/MSP や PureData）を用いることが圧倒的に多い。これはコンピュータ音楽の制作者がプログラミング言語やプログラミングそのものに精通しているとは限らないことに起因しており、比較的容易にプログラミングできるビジュアルプログラミング環境を当面の手段として選択することになるためである。また、インタラクティブなソフトウェアを作成する場合には、タイミングや音量バランスなど様々な処理のレスポンスを試行錯誤しながらプログラミングする必要があり、インタプリタ的に動作させながらプログラミングできるという利点も大きい。ただ、Max/MSP や PureData はプログラミングの簡便性やリアルタイム性を追求するあまり、音響信号処理の計算精度や処理速度の事情から音質がほかの環境ほど良くないことが指摘されている（ここでの音質の違いは、比較的高品位のヘッドフォンやスピーカで聴く必要がある）。その中で同様のプログラミング環境をもつ Kyma X だけは DSP 搭載の専用ハードウェアを用いるため、リアルタイムで高品質・高精細な処理が可能で音質の面では問題がない。なお、Max/MSP については高精細な音のレンダリングを行うための非リアルタイムモードが備わっているほか、Max5 では多少は改善されている。

一方で、PC の計算能力の向上により、リアルタイムでも高品質・高精細な音響信号処理ができる環境として Csound や RTcmix などが挙げられる。これらはテキストプログラミングや独特のデータ記述フォーマットを必要とするため、それらの記述や実行に不慣れな人には敷居が高いが、ビジュアルプログラミング環境では実現不可能もしくは実装困難な処理が、用意された機能やライブラリ群によって容易にできる点が特徴である。特に緻密な計算による音響を用いたコンピュータ音楽作品を追求する作曲家や制作者にはこれらの環境がよく利用される。

それ以外のトピックとしては、最近ではサウンドや音楽をプログラミングしながら観客に聴かせる「ライブコーディング」と呼ばれる行為が注目されており、その用途で比較的良好に利用される環境としては、SuperCollider, ChucK が挙げられる。SuperCollider は SmallTalk ベースの完全なオブジェクト指向言語であるため、一般的には独特なプログラミング及び実行環境としてとらえられるが、SmallTalk に精通していれば複雑な処理が効率良く記述でき、プログラムの再利用も容易であるなど、コンピュータ音楽制作環境としては奥が深いものとなっている。また、ChucK については音楽・音響処理の軽量プログラミング言語として脚光を浴びており、手軽に音響や音楽的な処理をプログラミングする環境として注目されている。

以下、これら音楽・音響を生成するプログラミング言語・環境の主要なものについて個別に説明する。

1-11-1 Csound^{1,2)}

MUSIC V などの MUSIC N 系の流れを汲み、Barry Vercoe によって開発されたリアルタイム合成可能な音響合成用言語である。現在の Csound5 は LGPL のオープンソースソフトウェアとして公開されている。合成には音響情報の orchestra ファイルと、楽譜情報の score ファイルの記述が必要だが、それら内容を XML として一つにまとめた csd ファイル形式でも記述できる。また、これらの合成はコマンドベースで行えるほか、C/C++ の API が用意されている。これらの API や Tcl や Python にもバインディングされているので、スクリプト言語でも音響処理アプリケーションが作成できる。Csound の合成エンジンの技術は AnalogDevices 社が DSP チップ化しており、いくつかの音響機器に搭載されているほか、MPEG-4 の SAOL の基盤技術として採用されている。

1-11-2 RTcmix³⁾

リアルタイム音響合成・信号処理のための C/C++ ライブラリ。サンプル単位のスケジューリングがロバストである点などリアルタイム処理の正確さが最大の特徴となっている。1980 年代前半に IBM メインフレーム向けのサウンドファイル入出力や変換を行うためのプログラム CMIX が開発され、様々な改良が加えられてリアルタイム処理向けとなったものである。数多くの信号処理や合成処理の機能をもっており、物理モデルによる合成エンジンには、Perry Cook と Gary Scavone による STK (Synthesis ToolKit) を採用している。元々は C 言語ライクなデータ記述言語 Minc を用いてスコアファイルを記述し、それをコマンドラインで操作するプログラムである。それ以外の利用手段として C/C++ でのプログラミングが行えるほか、スクリプト言語の Python や Perl のインタフェースもあり、プログラミング環境の選択肢はいくつもある。

1-11-3 ChucK⁴⁾

Ge Wang の博士論文として研究された音楽・音響処理のための軽量プログラミング言語である。ユーザは音や音楽の表現に注力できることを念頭においた環境がコンセプトとしており、ビジュアルプログラミング環境のような実行しつつ調整できるような処理環境を目指したものとなっている。そのために内部処理ではタイミング処理や並列処理に重点が置かれており、音響合成や分析処理は安定して動作するためライブコーディングなどでもよく利用さ

れる。文法としては、ChucK 演算子 (\Rightarrow) など独特の記述もあるが、軽量プログラミング言語らしい比較的的理解しやすい仕様であり、手軽に取り組めるものとなっている。

1-11-4 SuperCollider^{5,6)}

James McCartney が開発したリアルタイム音響信号処理・音楽記述言語で、SuperCollider3 以降はオープンソースとなっている。Smalltalk ベースのオブジェクト指向言語で、MIDI や信号処理のクラスのほか、GUI クラスなども用意されており、SuperCollider のみで GUI の音楽アプリケーションを作成することもできる。また、SuperCollider3 からはシンセシスサーバと言語処理クライアントとに分かれ、内部的に OSC (Open Sound Control) プロトコルで通信して処理を実行するアーキテクチャとなっている。そのため OSC が利用できるほかのプログラミング言語からも SuperCollider を制御することができる。

1-11-5 CLM (Common Lisp Music)⁷⁾

CLM は MUSIC V 系統の音響合成用言語であり、スタンフォード大 CCRMA の Bill Schottstaedt によって開発され、ソースコードのかたちで配布されている。その名の由来どおり元々は Common Lisp 向けのものだったが、Lisp 方言である Scheme でも利用できる。また、Lisp に限らず C 言語や Forth, Ruby でもプログラミングできるものとなっている。音響合成や信号処理はリアルタイム処理が可能のほか、グラフィックス表示機能なども備えており、CLM だけで音楽アプリケーションを作成することもできる。その例として、サウンドエディタ Snd は CLM で実装されたものである。

1-11-6 Haskore⁸⁾

関数型言語 Haskell の音楽記述用モジュール集である。音響信号処理を扱うためのものではなく、音符列に対して移調やテンポ変化を指定できるなど、音符や楽譜レベルの音楽処理記述を行うためのものである。処理結果は MIDI ファイルや Csound の score ファイルなどに変換される。

1-11-7 Max/MSP⁹⁾

音楽を扱うインタラクティブアートで世界的に使われているビジュアルプログラミング環境で、Cycling'74 社の製品。機能を割り当てたオブジェクトボックスをパッチコードで接続してゆき、データフローを記述する形式でプログラミングを行う。当初 Max は MIDI 処理向け言語として Miller Puckette によって開発され、David Zicarelli によってビジュアルプログラミング環境へと発展した。後に Pure Data の音響信号処理オブジェクト集 (プラグイン集) である MSP が登場し、リアルタイム音響処理も可能となった。最近では CG 処理のオブジェクト集が登場し (Jitter, GEM, DIPS など)、音楽もグラフィックスも同様に扱えるマルチメディア処理用プログラミング環境としても発展している。オブジェクトは C や Java で作成でき、世界中のユーザが追加機能を開発・公開している。

1-11-8 Pure Data (Pd)^{10,11)}

Miller Puckette が開発したリアルタイム音響処理機能をもつビジュアルプログラミング環

境で、Pd と略される。NeXT/ISPW 向けの Max (Max/FTS) のアーキテクチャを見直して、一から設計・開発されたもので、修正 BSD ライセンスのオープンソースソフトウェアである。Max とほぼ同様の記述・動作が可能で、マルチプラットフォームな環境であることから、Windows や Linux 環境のユーザを中心に普及している。Max/MSP の MSP は Pd の音響信号処理部分をベースに開発されている。

1-11-9 OpenMusic¹²⁾

OpenMusic はフランスの IRCAM で開発された音楽用ビジュアルプログラミング環境であり、GPL ライセンスのオープンソースソフトウェアとして提供されている。Max/MSP のようにオブジェクトをパッチコードで接続する形式のプログラミングを行うが、オブジェクトはアイコンのほか波形や五線譜などグラフィカルに表現されたものが多用されるため、プログラムの見た目は Max/MSP よりも直感的であるといえる。内部処理は、Common Lisp による CLOS (Common Lisp Object System) で記述されている。クラスやライブラリが多数提供されており、様々な音楽制作に対応可能である。

1-11-10 Kyma X¹³⁾

Symbolic Sound 社による音響合成環境。複数の DSP が搭載されたハードウェアサウンド処理エンジン Capybara を PC と FireWire で接続して、PC 上で動作するビジュアルプログラミング環境 Kyma で制御する。Kyma では、様々な音響信号処理モジュールがアイコンとして表現されており、パッチコードで接続して処理を記述する。ほかに波形エディタやタイムラインエディタなども Kyma には整備されており、多彩な音楽制作環境として機能する。

1-11-11 MUSIC V

Max Mathews が開発した音響合成用コンパイラであり、Csound や CLM の音響合成エンジンとして利用されているなど、世界的に広く普及している。

1-11-12 MusicXML

MusicXML は五線譜の音楽情報を様々な形式に変換して扱えるようにするための楽譜データ用 XML である。これ自体は音楽生成や演奏のための言語ではないが、様々な音楽ソフトウェア間でデータを橋渡しする用途に有用である。

1-11-13 SAOL

SAOL は MPEG-4 規格における音響合成・エフェクト処理のための記述言語であり、Csound を拡張した仕様となっている。合成用の記述やエフェクト処理の記述をデータとして通信し、受信側で様々な音響合成や波形データにエフェクト処理を実行することができる。MP3 などよりも 1/10 程度のデータ通信量で済むとされている。

■参考サイト：

- 1) <http://www.csounds.com/>
- 2) <http://sourceforge.net/projects/csound>

- 3) <http://rtcmix.org/>
- 4) <http://chuck.cs.princeton.edu/>
- 5) <http://sourceforge.net/projects/supercollider/>
- 6) <http://supercollider.jp/>
- 7) <http://ccrma.stanford.edu/software/clm/>
- 8) <http://www.haskell.org/haskore/>
- 9) <http://cycling74.com/products/maxmspjitter/>
- 10) <http://www-crca.ucsd.edu/~msp/software.html>
- 11) <http://puredata.info/>
- 12) <http://recherche.ircam.fr/equipes/repmus/OpenMusic/>
- 13) <http://www.symbolicsound.com/>