

音楽認知研究の諸問題

平賀 譲

図書館情報大学

計算的アプローチによる音楽認知研究では、音楽認知をどのような過程としてとらえ、それに基づいてどのような問題設定を行い、何を実現すべきかについて、明確な認識を持つことが重要である。認知という観点から眺めた場合、通常の音楽分析などで扱われる概念そのものの見直しが必要となることも少なくない。本稿では音楽認知研究で扱われる課題について、批判的見地から検討を行う。取り上げるトピックは、楽譜レベルへの記号化、調性・拍節構造、グループへの分節、階層構造、類似性の認識などである。全体を通じて、協調的・競合的分散処理としての音楽認知観を示す。

Issues in the Computational Study of Music Cognition

Yuzuru Hiraga

Univ. of Library and Information Science

1-2 Kasuga, Tsukuba, 305, Japan

hiraga@ulis.ac.jp

In a computational approach to music cognition, it is essential to hold a clear, rigid view of how music cognition is to be characterized, what problems are to be pursued, and what should be realized. From a cognitive viewpoint, it is often the case that the concepts incorporated in traditional styles of music analysis must themselves be totally revised. This paper presents a critical introduction to the issues raised in music cognition research. The discussed topics include score-level symbolization, tonal and metric structures, group segmentation, hierarchical structures, and detection of conformance. The presented view is based on the recognition of music cognition as cooperative-competitive distributed processing.

1 はじめに

本稿は、根底の意図としては別に記した原稿¹と同じ内容を扱ったものである。ただしここではよりストレートに、またざっくばらんに記していきたい。

音楽認知に対する計算的アプローチ、つまり認知の計算モデルを作成しようというアプローチは、二重の意味でアイデンティティを確立する必要がある。1つは音楽認知研究全般の中においてである。音楽認知に限らないが、計算的アプローチの根本目標は「脱意味化 (desemanticizing)」にあるだろう。日常的な音楽体験、音楽心理実験、音楽理論などから得られる知見は、理論的にも情緒的にも「意味豊かな」世界の記述である。その本質的な部分を、抽象化された情報に対する形式的な操作の体系として理解しようというのが脱意味化の意図である（物理学が、アニミズムや神学による説明を数理的な記述に置き換えたことになぞらえられる）。残念ながら現在の研究水準は、そう見なしうるレベルには達していない。つまり独自のまた先導的な概念的枠組を提供するというよりは、他のアプローチにおける知見への後追い確認的な面の方が強い。

もう1つは、ほかならぬ音楽情報処理一般の中において、特に問題となるのは（計算機を用いた）音楽分析との関連である。人間の音楽認知がそのような分析をそのまま実行しているというなら話は簡単で、音楽分析すなわち音楽認知の研究ということになる。そのような立場をとる研究者も少なくないが、これは楽天的にすぎる。もちろんオーバーラップする部分は存在するにしても、音楽分析はあくまで「音楽について」の理論であって、我々の認知過程がそれに対してどう働くかを直接教えてくれるわけではない。音楽理論が提示する概念を、普通人が（たとえ意識下にせよ）遂行している保証はないし、逆にそれらでは看過されていても、認知の観点からは本質的な要因もあろう。したがって問題は両者の切り分けをどう行うか、特に「認知」という観点をどのように確立し、それを具体的研究に反映させるかである。これには「有用性」といった問題も絡んでくるので話がややこしくなる。

以下ではそういった話題を取り上げていきたいが、ざっくばらんということで、本格的な論陣を張るのではなく、むしろ問題提起として具体的なトピックをいくつか取り上げ、議論のタネとしたいと思う。

2 楽譜、書けますか

楽譜が音楽の抽象的表現であることはまちがいないし、「採譜」が音楽情報処理にとって重要なものも明らか

である²。認知においても楽譜レベルの抽象化——いわばメンタル楽譜の作成は、当然のことと想定されている。しかし本当にそうだろうか。

実際に楽譜が書けることは、そのような楽譜レベル処理の（一応の）証拠にはなる。Longuet-Higgins[12]などは、はっきりこれを認知能力のうちに数えている。しかし言うまでもなく、採譜というのは結構特殊な技能であるし、ついでながら、込み入った曲になると専門家でも元の楽譜を再現することは事実上不可能である。それどころか、普通人だと絶対音感はおろか、相対音感（ある調性のもとで、個々の音の音度を特定する）でさえおぼつかない。つまり音を個別に同定することはできない。少なくとも、孤立した音の羅列という意味での楽譜表現が、認知的に実在しているかは疑問になる。

もちろんこれに対し、それらの処理は意識下で行われており、明示化・言語化できる必要はないという反論はできる。しかしこれは「意識下」に問題を押し込めただけで、我々が明らかに、しかも意識的に曲を認識し、記憶できることとのギャップは残る。

曲には覚えやすいものもあれば覚えにくいものもある。これは楽曲の記憶が個々の音符によってではなく、それらによるパターンとして構造化されていることを強く示唆している。また同一あるいは類似した曲の照合や、メンタルな変換（長調 → 短調の移調など）も、構造的表現の上での操作と考えた方が自然である。その一方、知っている曲を思い起こすとき、我々はそれを頭の中で「演奏」してみる。これ自身興味深い、それはさておき、このメンタル演奏はかなり具体的なイメージ——演奏する楽器とか歌手の声質、時には特定の演奏とさえ結びついている。いうなれば「エピソード記憶」であって、楽譜のような無機的・中性的なものではない。奇妙なことに音楽の記憶は、このように極めて具体的か、逆に抽象化・構造化されているかの両端に分極していても、その間の楽譜レベルに相当する部分は抜けているようである。

ここで指摘したいのは、音楽を CDE... とか Imaj7 のように、記号化して扱うことの危険性である。これには2つの側面がある。1つは認知システム内でそのような記号表現が、文字通りに存在するとは考えにくいことで、これは上に述べた。もう1つは研究方略上の問題で、このような記号表現をシステムへの入力として用いるのはいいいとしても、音楽的な「意味」を生み出す要因はすべてシステム内で定義しなければならない。ヒルベルト・プログラムを持ち出すまでもなく、記号表現系では個々の記号自体は無意味（無定義語）であって、「意味」は記号どうしの関係（やその上の操作）として存在する。これ自体は計算的アプローチではある程度まで不可避だ

¹ 情報処理学会誌特集「音楽情報処理」：「音楽認知（研究）はなぜ難しいか」

² ここでは楽譜に書ける「普通の」音楽（主に西洋調性音楽）を考える。

が、たとえば「音度は主音からの音程差で求められる」といった表面的な処理では話は済みそうもない。真の脱意味化には、音どうしの関係の膨大なネットワークが必要になる。

この記号化の問題は和声の場合に特によく当てはまる。和声学では和音は Cmaj7, Ddim のように記号化され、和声規則はそれらの記号の間の関係を指定するという形で定式化される。人間にとって和音は明らかに認知的な「質感 (qualia)」を持っているが、これ自体は記号化された規則には反映されない。しかも和声学は本質において帰納的な理論であり、それが対象としているジャンルには適用できてもそれに当てはまらない曲種には無力である。したがって人間の和声認識、より一般的には協和性の認識に対する原理的な説明力を持たない。このような知識は人工知能で言う「浅い知識」あるいは井口 [8] の言う「なぞり知識」にすぎない。利用目的がはっきりした工学的立場ではこれが許容されるし、推奨さえされる。しかし認知研究としては、より原理的で robust な探究が必要である。

単音がそれ自身では無性格であるにも関わらず、それらが組み合わされると認知的な質感が生じるというのは、考えてみれば不思議なことである（これは和音だけでなく、継時的なつながりも含めてである）。言語で言えば文字と単語の関係になぞらえられるが、言語の場合は両者の結びつきが恣意的であるのに対し、音楽ではその結びつきに少なからぬ必然性が感じられる。その理由が倍音構成など、物理的性質にどこまで帰着できるか、どのような生理的メカニズムによるのか、どこまでが生得的で何が獲得されるかなど、問題は多い。現在のところ、いろいろな試みはあるものの、計算的立場からこういった音の協和性について、十分満足のいく理論は存在しないと行ってよい。そしてこれが音楽認知研究の最大の課題であり、またネックになっていると言えるだろう。

3 調性：局所と大域

調性は構造的認知のための最も基本的な枠組を与えるものである。「ハ短調交響曲」などと言う場合、曲全体（楽章等）が1つの調に支配されていることを指す。調は旋法（長調、短調など）と主音の2つの属性に分解できる。旋法は物差しが目盛りの振り方に、主音は原点の位置決めに対応する。曲の認識は主音の取り方によらないと認められるから（これもいろいろ問題はあがる）、本質的なのは旋法の方である。

曲が単一の調を持つという前提のもとでの調決定には、記号レベルで見てもいろいろな手がかりがある。たとえば最終音を主音とするという「規則」だけでも半分以上の正解率にはなろうし、音高の頻度分布から最も適合度の高い調を選ぶ方法もある。和声があれば、和音はそれ自身で調性感に対応づけられるし、カデンツバター

ンなども参照すればより正確な判断ができる。もちろん実際にはそう単純ではない。簡単な曲でも調が途中で変わる、つまり転調するのが普通だからである。その場合、調は曲の部分部分について局所的に決まることになる。そこで転調（感）が生じるのはどういう条件によるかが問題になる。転調部分の長さをとると、フレーズのようにある程度以上の長さが必要という可能性もある一方、極端に言えば、1つの和音（たとえば属和音）でも局所的な転調感を生じるという見方でもできる（借用和音などはもう少し現実的なケースである）。

人間の認知において、調感急速に確保されるし、瞬時に変更もされる。つまり調性は局所的な概念であり、その意味において不安定でもある。曲が単一の調、つまり大域調を持つというのは、音楽（作曲家）がそういった認知特性に合わせ、安定した曲認識を可能にしようとした「結果」にすぎない。局所調という観点からは、大域調の場合とは違った認識が生まれる。まず唯一の調、「正しい」調を決めなければならないという強迫はなくなる。しかし一方、局所的なデータで判断することになるので、曖昧、不正確になる。さらに大域調との折り合いをどうつけるかも問題である。

実時間で調決定を行おうとすると、必然的に局所調の立場に立つことになる。これは音楽認知だけでなく、（むしろ）実時間合奏などにも重要である。Krumhansl [9] は心理実験に基づいて図1のような音階音の「プロフィール」を導出し、これによって音高の頻度集計に対する「重みづけ」を行い、すべての調の可能性から適合度の高いものを選ぶという調決定アルゴリズムを与えた。これは曲頭の数音からでも、「（大域調として）正しい」調を高い正解率で選択できる。処理対象とする音を適当なタイムスライス中に限定すれば、そのまま局所調の判定にも使える。これは統計的手法を用いているが、Bharucha [2] は同様の処理をニューラルネットで実現している。他にも Krumhansl が比較対象としている Longuet-Higgins と Steedman [12] のプログラムや、和声に基づいた Rowe の Cypher [19] などがある。

Krumhansl の方法自体は簡単なものであり、いろいろな問題がある（たとえば音高頻度だけを用いているため、音列としての構造が扱われていない、タイムスライスは手作業で設定しているが、区切り方自体が後述のグループ化とも関連して問題であるなど）。これについては著者本人も詳細な考察を行っており、ここでは触れない。Krumhansl によれば、専門家でも局所調の判定については意見の相違が生じる（そもそも局所調という概念自体、従来の音楽学にはないと言うべきか）。しかしむしろ多くの場合、特に単旋律では、局所調は客観的には曖昧と見る方が正しいだろう。阿部 [1] は図2のような音列から「終始音導出」という方法で、調性を判定させる実験を行った。図2の場合、F, C, A などが終始

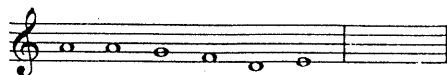


図2：終始音導出実験の音列例 [1]。

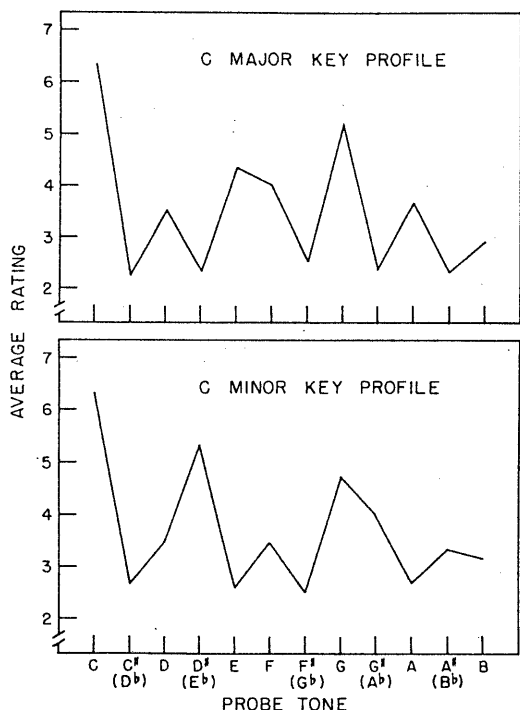


図1：音階に対する12半音の心理プロフィール [9]。

音（主音としてよい）に選ばれたが、他にも可能性はある。これは調判定が単なる客観的な処理結果というよりは、先行文脈 (priming) や聴取者自身の「構え」も重要な要因であることを示している。たとえば Beethoven の「運命」交響曲の冒頭、例の“GGGEb”という音列は、Krumhansl 等のアルゴリズムでは、ハ短調よりはむしろ変ホ長調が優先されてしまう（続く“FFFD”を加えてもあまり事情は変わらない）。にも関わらず、我々がまったく曖昧性なくハ短調を感じるのには、単に曲の知識があるからか、あるいは別の要因によるのかはともよくわからない。

局所調判定では可能な調すべてに対し数値的評価を与え、値の高いものを選ぶという方法をとることが多い。阿部も述べているように [1]、そのような全列挙的な処理が認知的に妥当性があるかは疑わしい。また旋法、主音の属性が区別されずに処理されるが、主音がわからなくても旋法についての質感はある（あるいはその逆）という意味で、これらは独立性があると考えられる。たとえば短調における自然短音階や和声的短音階、あるいは日本音階の短旋法などは、質感として共通の性質を持っていることは明らかだが、そういった旋法間の近縁関係などを、計算的枠組でどう定式化すればよいかは見当がつかない。もっともこれは協和性の問題に踏み込んだ話ではある。

最後に大域調との関係に触れておこう。安定した大域調が認識されるとすれば、それについての記憶が保持されているはずである。曲自体がその調に留まっている場合には何も問題はない。局所調の認識がそのまま大域調の認識を持続させるからである。ごく部分的な転調などでも、大域調に対する「活性化」が残存しているとすればよい。ところが長時間にわたる転調となると、問題が生じてくる。そもそもそのような場合、曲が原調に戻った場合、「復帰」として認識されるのだろうか。音楽学ではこれが当然のことと受けとめられているようだが、筆者などは数分も転調が続くと、原調の記憶があるのか自信はない（もっとも曲自体の記憶は結構あるのだから、少なくとも旋法の記憶はあることになる）。逆に原調に復帰しなくても不自然でない曲は、Bach の「フーガの技法」などを上げるまでもなく、いくらでも存在する。しかし原調の記憶保持を認めるとすると、上記のような遷移的な局所調判定にどう反映されるかは問題になる。原調に対する活性化が残るという考えではうまくいきそうにない。局所的な調自体の曖昧性（原調か転調された

調か)と区別することができないからである。もし記憶が保持されるとするなら、別の機構を考える必要がある。その場合でも、複数の調にわたる場合、どれについての記憶がどう残るのかは、いずれにしても問題である。

4 拍節：諸要因の交錯

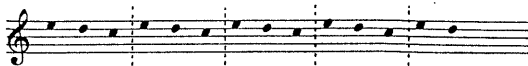
小節や拍 (beat, tactus) など、周期的な時間区切りを与える構造を拍節構造 (metrical structure) と呼ぼう。拍節についても調性と同様の局所一大域の問題が存在するが、ここでは別の点を取り上げる。

実際の演奏と楽譜が示す拍節の長さなどが一致しないことはよく知られている。そこで楽譜レベル表現を得るには音長の正規化ないし量子化 (quantization) が重要課題となる [3, 11, 12]。これは楽譜が「正しく」、演奏はそれからの逸脱という立場である。採譜が目的ならそれで問題はないが、ここでも発想を逆転して、(個々の)演奏の方が「正しく」、楽譜はその近似にすぎないという見方もできる。なぜ近似かは簡単で、そこまで書ききれないからである (演奏には個性的側面もあるが、基本的なパターンには共通性のあることを注意しておこう)。こちらの立場では拍節はより定性的、少なくともフuzzyな概念になる。

なぜ拍節が存在するかについてはいろいろな詮索ができる。人間はもともと規則的なリズムに強い反応を示すものだが、それだけが理由ではないだろう (拍節を持たない音楽もいろいろ存在する)。構造認知の観点から重要な要因として、拍節が確保されていればグループへの分節が効率よく行えること (短い時間の中に押し込めても区切りがはっきりすること)、曲全体の統合を崩すことなく、自由で複雑な展開ができることなどがあげられる。しかしメトロノームを用いて演奏するのもでない限り、拍節自体も曲自体から認識されるものであり、それにはグループとしての区切りが主要な手がかりになる。つまり拍節の認識とグループの認識とは、現象としては独立でも、処理としては相互依存関係にある。この見方は上の定性的立場と符合する。また異なるレベルのグループが異なるレベルの拍節に直接対応するという意味で、一種の多重時間構造が想定できる [4]。これは拍節認識がボトムアップに階層的に構成されるという立場とは対立する見方である。

拍節抽出 (beat tracking) の研究の多くは、音長情報だけを利用している [3, 11, 12]。これは実はやや奇妙な考え方で、拍節の強拍 (小節や拍の開始位置) にはアタックが存在する、逆に言えば、アタックのない位置は強拍ではないという「不在の検出」に基づいていることになる (拍が「強い」というのは構造上の概念で、音が文字通り強く演奏されるとは限らない。実際、通念に反して、アクセントや奏法は、拍節の手がかりとして必ずしも有効ではない [12, 18])。当たり前のことだが、これでは

(a) 3 拍子の群化パターン



(b) 4 拍子の群化パターン



図3：拍節認識の2つのパターン [1]。

等拍リズムには無力である。

音高情報なども拍節の認識に明らかに寄与する。たとえば図3 [1] で、(a) は音高パターンから3つ組 (3拍子) として認識するのが自然だろう。しかし文脈や聴取者の「構え」によっては、(b) のように2拍子系の認識が優先されることもある。拍節とグループ化が競合する具体例としては *Moldau* の冒頭や、Chopin の曲の多く (たとえば「黒鍵のエチュード」の最後)、極端な場合として「夢窓花」のような hemiola がある。

このように拍節構造1つとっても、音の時間パターンだけでなく、様々な要因が交錯する総合的な認識である。これ自体は別に目新しい話ではないが、いざ計算的アプローチで実現するとなるとたいへん難しい。要因間の相互作用が組み合わせ論的に効いてくるし、その調整も困難だからである。ただ次のことは注意しておこう。認知システムがモジュラー、つまり単機能に分解可能であるというのは、(方法論的には) 当然の仮定である。音長情報だけを用いた拍節抽出にしても、そのようなモジュールの1つと見なすことができる。それらモジュールは、単独では必ずしも十分な判断ができるとは限らない。そこでそれ自身としての機能を詳細に検討することが重要である。特に何が「できないか」、どういう場合に失敗するかは、成功するケースを増やす以上に大切になる。特定の機能を通じて見ることににより、音楽の持つ曖昧性、したがって構造の複雑さが明らかになり、どのような相互作用があるかを探ることもつながるからである。逆に処理結果の「正解率」だけをアドホックに追究するのは、あまり意味はないことになる。

5 グループ化の落とし穴

音楽が部分的な単位 (グループ) に分節できることは「楽式論」のテーマであるし、Lerdahl と Jackendoff の理論 ([10]: 以下 *LJ*) でも「グループ構造 (grouping structure)」として柱の1つになっている。*LJ* でグループを定義する構成規則 (well-formedness rules) は、次のように要約できる。

グループは時間的に連結した曲の部分であり、厳格な階層構造により構成される。つまり曲全体が最上位のグループであり、上位グループは下位グループ群に、隙間も重複もなく分割される (局所的な重複には例外として許さ

れるものがある)。

グループ化自体はごく自然な概念であり、それを疑う余地も必要も感じない。しかし具体的内容となると、いろいろな疑問が生じる。*LJ*ではグループを具体的に決定する要因を選好規則 (preference rules) として与えている。これらは「望ましい」グループ化のための条件を規定するもので、それを最大限満たすグループ構造が選択される。ただちに生じる疑問は、グループ化が一意に決定できるかである。選好規則は各要因がどう寄与するかは規定するが、それらの相互作用を示すことは意図的に回避され、総合判断は分析者 (あるいは認知システム) に委ねられている。当然、要因間で不整合が生じるケースはある。実際、専門家の間でも分析が食い違うことはよく知られている。

真の問題は、この不整合をどうとらえるかにある。仮にすべての要因が一致すれば明らかな分節が感じられるだろうが、音楽はそこでストップしてしまうだろう。すると不整合というのは、グループへの分節に曖昧さを生じるという消極的な意味ではなく、分節が起こってもそれにまたがって曲の連続性を保つという積極的な意味を持つことになる。筆者も試みたことがあるが、厳格な階層的グループ化ではどうしてもこのつなぎ情報が欠落してしまう。実際に聴いた印象では顕著であるにも関わらずである。これ自身もグループとして認めてしまうと、もはやグループは曲を整然と分割するものではなく、互いにオーバーラップして絡み合う存在となる。これは階層的グループ化自体を否定するものではないにしても、その重要性を相対的に低いものにはする。その点 *LJ* がせっかく選好規則という形をとりながら、なぜ一意的なグループ化に執着するのかはよくわからない。

*LJ*のグループ化規則は、ゲシュタルト心理学のグループ化原理に依拠している。たとえば音程が近い音どうしは離れたものよりグループとして結びつきやすいという規則などである。これによれば、同じ音度は最も近い音程として、最も強く結びつくことになる。しかし同音については別の見方もできる。強い結びつきが必要なら、つの音にまとめてしまえばいいわけで、わざわざ分けるからには、それなりの理由があるはずである。そのような可能性のうちで最も極端なのは、そこでどうしても分節させたいという場合である。本来つながるべき同音が分割されることで、強い構造的ギャップを指示するわけである (拍節の境界でシンコーションを避けるのは一番簡単なケースである)。つまり同音の連続は、ある場合には強い接続、ある場合には強い分節という、相反する機能を果たしうることになる。

一方、上の規則では広い音程は構造的関係が希薄と受け取られることになるが、これも「広い音程」という「関係」は存在するのであって、真空というわけではない。一般化すると、どのような音程にも「関係」は存在

するのであり、その関係 (順次進行、アルペジオなど) が持続されることこそがグループ化の基礎と考えるべきである。たとえば前後の順次進行の間に同音連続があれば、そこだけ異質の関係として分節が示唆されることになる。*LJ*では個々の音の属性のみをとらえ、一段高次の音どうしの関係に目を向けておらず、これは大きな問題点となっている (ゲシュタルト原理の good continuation が言及されていないのは象徴的である)。

6 一様な階層性

*LJ*のもう1つの柱は「簡約構造 (reduction)」である。簡約とはグループ中から代表となる1つの要素 (単音や和音) を選択することを指す。それら構造音の相互関係としてどちらが重要かを規定することにより、グループどうしが有機的に結合されることになる。この簡約操作を再帰的に適用することにより、階層的な簡約構造が得られる。簡単に言えば、装飾的・副次的な要素を次々に取り除いていった、曲の骨格に相当する。*LJ*には2種類の簡約構造があり、1つは隣接するグループどうしの構造的な重要性に基づく簡約、もう1つは調性的な緊張-弛緩関係に基づく簡約である。

グループ構造の場合でもそうだが、これらの階層構造は再帰的、つまり全体として一様な構造をしている。上位レベルでも下位レベルでも、基本的には同じ構成原理に基づいているわけである。しかし特に大規模な曲になるほど、この一様性が適用できるかは問題になってくる。たとえばソナタ形式で、提示部と再現部の関係づけはどこまで可能なのだろうか。あるいは曲全体が緊張から弛緩 (解決) へ向かうと述べることにどれだけ意味があるのだろうか。むしろ簡約が結果として切り捨てるものの一つになればメロディラインやリズムのような曲の微細構造 — の方に、認知的な顕著さ (salience) があるのではないだろうか。

言語の場合で考えると、文の中での句・節といった階層は、(生成規則で記述できるという意味で) 一様と認められないことはない。しかし文どうしの間の「文法」——いわゆる談話文法——は明らかに文レベルの文法とは異質のものである。音楽の場合も後者のような、異質な階層性を持つと考えられる。大域的な構造は、文の羅列のような平坦で線型なものか、あるいはもう少し発展させて、前述のような (グループどうしが交錯しあう) ネットワーク的なものと考えの方が直感的なイメージに合う (Narmour [17] も同様の指摘を行っている)。大規模で深度の深い階層構造は、単に理論的整合性のための artifact にすぎないという気がしてならない (最初に述べたように、我々は音楽がどう作られているかではなく、どう聴かれるかを考えたいのである)。

7 構成原理としての類似性

音楽、特に西洋調性音楽では、同一ないし類似したパターンの繰り返しが基本的な語法になっている。たとえば AA'BA という楽式では、パターン A が提示され、その変形が続き、対照的なパターン B のあと、再び A が示される。情報理論的には明らかに冗長な情報で、これが音楽で多用されていることは興味深い。

その重要性にも関わらず、音楽理論で類似性や繰り返しを論じているものは驚くほど少ない。LJ でも重要な構成要因として強調されているが、それ自体についての議論は行わないと断られている。正面切って論じているものとしては、わずかに Meyer[13] が上げられるくらいである。むしろ計算的アプローチの方で、パターン認識の観点から取り上げた研究が多い [5, 6, 20]。

類似性の問題は、どのような要因のもとで類似性が成立するかと、認知システムの中でどのように処理され、どういう役割を果たしているかという問題とに分けられる。前者について実際の曲を調べてみると、類似性のパターンがかなり多様であることに気づく [6]。平行移動のように系統的な変換もあれば、部分的な変化によるものもあるし、音程、リズムなどの特徴的な部分を抽出・強調するものなど、様々である。前節の簡約も類似性の一種にはかならない（簡約の逆操作は装飾や変奏になる）。これらはある程度までは形式的な照合操作で発見できるが、人間の直感的な認識まではなかなか到達できない（念のため——学生研究などに見られる機械的な相関手法による検出では、類似性のごく特殊なケースしか扱えない）。

これが後者の機能・役割の方になると、極めて厄介な問題になる。仮に類似性照合プログラムがあるとして、どこに類似性が存在するかの手がかりがないと、ひっきりなしに照合を起動する状況になってしまう。これは計算量もバカにならないが、さらに悪いことに、直感的には類似が感じられないゴミまで大量に拾ってきてしまう。これはプログラムのまずさばかりのせいとは言えない。類似性を過剰に見いだそうとする理論家 (Reti) に対する批判が Meyer[13] にあるからである。

類似性が難しいのは、単にそれを発見すれば済む話ではないからである。類似性自体はグループを単位に成り立つものであろうが、逆に類似性の発見がグループへの分節の重要な、というより決定的な手がかりになる。したがって類似性の検出自体、グループの認識に先行ないし並行している必要がある。ここでもモジュールの相互作用の問題が浮上してくるわけである。

類似性は曲中の照合だけでなく、曲の再認、つまり既知の曲との照合にも関わっている。聴いている曲を識別したり他の曲に似ていると感じるのは、記憶との類似性照合が行われていることにほかならない。これも経験的には急速に処理され、客観的には判定できないと思われる時点でも、結果が認識されるようである。先ほどの「運

命」の例でも冒頭の動機だけで、他の曲の可能性（たとえば長調として続くなど）は認識される余地がなさそうに思える。このような記憶の表現形式や処理のメカニズムはおそらく極めて効率的な、data-driven な形態のものであろうが、実質何もわかっていないと言ってよい。逆に類似性の詳細な探究を進めていくことが、これを解明するための有用な手がかりになると思われる。

8 Narmour 理論と調性独立な処理

ここまで述べてきたスタンスは、基本的には Narmour による「暗意-実現モデル (Implication-Realization model)」の考え方と符合している。実際、筆者としては、Narmour の路線が音楽認知を理解する最も適切な方向を示しており、計算的扱いとも整合性があると考えている。ただ、Narmour 理論は膨大なもので、ここではその 1 つの側面 (Parametric System) だけを述べるにとどめる (Narmour 理論については [15, 16] を参照。これは [14] とはいくぶん異なる体裁になっている。ダイジェストとしては、たとえば [17] 参照)。

音楽認知をモジュラーな協調・競合の分散システムとして見る立場をここまでとってきた。モジュールとして調性処理を司るものが存在するなら、一方で調性に依存せずに構造解析を行うモジュールの存在も想定できる (局所調の判断が非確定的にしかできないなら、そのようなモジュールの重要性は高くなる)。このモジュールにとっては調性情報は存在しないから、音程情報として利用できるのは、上行か下行か、音程が同じか狭いか広いといった定性情報だけである (西洋音階以外の音階を見ても、順次進行をなす狭い音程として認識される幅には共通性があり (高々 3 度ぐらい)、しかも音階上で広い跳躍音程と混乱は生じない。この広い、狭いの認識は普遍的な、おそらく生得的なものと考えられる)。一般原則として、狭い音程はそれが継続される傾向を、広い音程は元の音へ戻ろう (逆行しよう) という傾向を持つ。これが実現されるか否かによって、旋律パターンは少数 (約 20) の基本的な定性パターンに分類される。この基本パターンの連結が旋律を特徴づけることになる。

筆者は Narmour 理論とは別に、このような定性的情報で曲の分節を試みている [7]。ただし、音長についても同様の定性的扱い (長い/短い) を用いた。本格的な検討はこれからだが、初期の印象として、これだけでもかなりの認識が可能である。このアプローチの大きな利点は、調性情報に依存しないため、旋法の違いやエラーなどに対して robust な処理が可能になる点にある。もちろん問題点はいろいろあり、その分析が最も重要な課題となる。特に前述のように、類似性の検出はかなり決定的な役割を果たすようである。

比喩的に言えば、上記は音楽を調性と、音程の相対運動という要因の直積としてとらえている。これらは単独

では(少なくとも意識の上では)認識しえないという点でクォークのようなものだが、このような要素分解という見方を通じて音楽認知の複雑さと robust さの一端なりとも理解できるのではと考えている。

9 おわりに

本稿では音楽認知をどうとらえるか、それをどのように研究していくかについて、著者の考えを述べた。特に重要なのは、特定の目的を指向した音楽分析に対し、人間の認知過程を解明するという立場をはっきり区別することで、その観点からいくつかの研究例を(主に批判的に)取り上げた。これはいたずらに問題設定を難しくしているだけだという意見もあるだろうが、逆にそうでなければ何を研究しているのかが曖昧になってしまう。むしろ、認知という観点から眺めることで、様々な音楽概念を見直すことになる点を強調したい。

ここで取り上げられなかった話題もいくつかあるが、それらについては別の機会に取り上げたい。

参考文献

- [1] 阿部純一:「旋律はいかに処理されるか」、in 波多野(編):「音楽と認知」, 東大出版会, pp.41-68 (1987).
- [2] Bharucha, J.J.: Music Cognition and Perceptual Facilitation: A Connectionist Framework, *Music Perception*, vol.5, pp.1-30 (1987).
- [3] Desain, P. and Honing, H.: The Quantization of Musical Time: A Connectionist Approach, *Computer Music J.*, Vol.13, No.3, pp.56-66 (1989).
- [4] Desain, P. and Honing, H.: Time Functions Function Best as Functions of Multiple Times, *Computer Music J.*, Vol.16, No.2, pp.17-34 (1992).
- [5] Deutsch, D. and Feroe, J.: The Internal Representation of Pitch Sequences in Tonal Music, *Psychological Review*, Vol.88, pp.503-522 (1981).
- [6] 平賀譲:「音楽認知のための知識表現」、in 波多野(編):「音楽と認知」, 東大出版会, pp.97-130 (1987).
- [7] Hiraga, Y.: A Computational Model of Music Cognition Based on Interacting Primitive Agents, *Proc. of ICMC 1993*, pp.292-295 (1993).
- [8] 井口征士:「音楽と感性」、in 井口他:感性情報処理, オーム社, pp.79-102 (1994).
- [9] Krumhansl, C.L.: *Cognitive Foundations of Musical Pitch*, Oxford Univ. Press (1990).
- [10] Lerdahl, F. and Jackendoff, R.: *A Generative Theory of Tonal Music*, MIT Press (1983).
- [11] Longuet-Higgins, H.C. and Lee, C.: The Perception of Musical Rhythms, *Perception*, Vol.11, pp.115-128 (1982).
- [12] Longuet-Higgins, H.C.: *Mental Processes — Studies in Cognitive Science*, MIT Press (1987).
- [13] Meyer, L.B.: *Explaining Music: Essays and Explorations*, Univ. of California Press (1973).
- [14] Narmour, E.: *Beyond Schenkerism — The Need for Alternatives in Music Analysis*, Univ. of Chicago Press (1977).
- [15] Narmour, E.: *The Analysis and Cognition of Basic Melodic Structures*, Univ. of Chicago Press (1990).
- [16] Narmour, E.: *The Analysis and Cognition of Melodic Complexity*, Univ. of Chicago Press (1992).
- [17] Narmour, E.: The Influence of Embodied Registral Motion on the Perception of Higher-Level Melodic Implication, in Jones, M.R., Holleran, S. (eds.): *Cognitive Bases of Musical Communication*, American Psychological Association, pp.69-90 (1992).
- [18] Rosenthal, D.: Emulation of Human Rhythm Perception, *Computer Music J.*, Vol.16, No.1, pp.64-76 (1992).
- [19] Rowe, R.: Machine Listening and Composing with Cypher, *Computer Music J.*, Vol.16, No.1, pp.43-63 (1992).
- [20] Simon, H. and Sumner, R.K.: Pattern in Music, in Kleinmuntz, B. (ed.): *Formal Representation of Human Judgment* (1968).

SIGMUS 第5回 研究発表会 質疑記録

1994年2月4日 NTT 武蔵野研究センタ

- (1) コード進行をベースにして作曲を支援するシステム
内橋由佳, 平野健太郎, 伊丹誠, 伊藤紘二 (東京理科大)
記録: 鈴木 (東京高専)

Q: 長嶋 (ASL) Sun から MIDI を出力する方法は?

A: インターフェースボードを自作している。

Q: コード中の基音以外のベース (分数コード) をサポートしているか?

A: 考えているが今のところはない。

Q: 1小節内の音符単位での分割はできるか?

A: できる。

Q: コードモーションの種類は?

A: 現在は C, F, G7 の3種類である。

Q: 次に移る隣のコードの動きだけを見ているのか?

A: そうである。2つ前は見ていない。

C: 小坂 (NTT) 今回はコード進行から見た作曲であったが、実際に使う立場から見るとメロディが先に浮かぶ場合が多いと思う。

Q: 高橋 コードネームに別の名前を付けられるか?

A: 違う名前も付けられる。

Q: 自作のコードでもモーションとして登録可か?

A: 登録可能である。

- (2) 標準 MIDI ファイルからのメロディの自動抽出法
鷺坂光一 (NTT)
記録: 増井 (富士通研)

Q: 長嶋 (ASL) (メロディチャンネルの選択例として用いた) 224 曲の MIDI ファイルデータの提供ベンダ数は? それぞれメーカーごとに癖がある。

A: ローランドとカワイなど.. 2, 3社である。

Q: ピアノソロ曲は1チャンネルにデータが全部入っているのか?

A: 1チャンネルに入っている。

Q: ソロ曲におけるメロディチャンネル選択結果の合否はどう判断するのか?

A: ソロ曲は 224 曲中には含まれない。処理内容の説明としてピアノソロを例にあげた。

Q: SMF には最初からパート毎にヘッダにメロディやコードといった記述が含まれているのでは? だとするとメロディ抽出の意義は?

A: ヘッダにおけるテキスト部分にはパート情報は必ずしも含まれない。用いた SMF データには含まれていなかった。

Q: 平賀 (国情大) 「1つのチャンネルを選択」とは「全曲を通じて」の意味か?

A: 歌謡曲を用い、歌詞の入っている声部が選択されたかどうかを判断した。

Q: 成功例では、途中でメロディを担当している声部 (チャンネル) が替わったりしないのか?

A: しない。

Q: (メロディチャンネル選択における) 失敗の原因は?

A: 音域のずれの問題, 伴奏の派手さなどが考えられる。

Q: 声部の交替が原因になって失敗しているものがあるのでは? 一般のデータでは、頻繁に声部の交替があるのが普通ではないか?

A: そうではなかった。

Q: 1つの声部だけがメロディを担当するというのは仮定として強過ぎるのではないか? 現実の演奏では何人もがメロディを担当することは多い。現実の音楽データベースとしての利用を考える場合、考慮が必要であろう。むしろ音楽データベースという観点からは、検索側で「メロディ」をきっちり押えておく必要があるのではないか?

A: 仮定が違う。メロディは意味空間を構成する一つの要素であり「意味検索」のためのツールと使う。メロディの指定によって音楽を検索するために使うのではない。

Q: 検索の Key はメロディではないのか?

A: 「意味検索」の Key は必ずしもメロディではない。オーサリングシステムにおいて絵にあった音を選ぶ場合に、「どういう絵か」「どういう音楽か」がそれぞれ定義できているならば、「その絵にふさわしい音」としてマッチングを取れる可能性がある。

Q: 鈴木 (東京高専) 「似たようなメロディの曲を探したい」というのが研究目的なのか?

A: 部分集合である。絵と音とのマッピングを行なうソフトがあるが、例えば、あるシーンで、絵と音のマッピング状況など、絵が持つ性質にしたがってサーチが可能かもしれない。

Q: MIDI は音楽情報としては記号化されたものである。データベースとして考えると、もっと一般的な PCM データを扱えると応用が広がるのでは?

A: 同感であるが、今回はまず MIDI データを利用した。

Q: 伊藤 (東京理科大) メロディ・コードの雰囲気と絵の雰囲気に関連されるやり方については?

A: 現在、マッピングについていいアイデアを探している。

Q: 自動化というより、人間が介入するのがいいのか? 絵の評価と音の評価をつなぐ感性的なチャンネルというのは、人間の感覚を介する以外にはないのか?

A: 例えば、悲しい絵に悲しい曲を結び付けるのは感性情報のレベルであるが..

Q: そのカテゴリーは人間がやるしかないのか?

A: 例えば人間と機械で異なる観点が存在するとして、その判断基準が存在すれば、そのマッピングによっては整合性がとれるかもしれない。

Q: その規範は人間がきめるべきなのか?

A: 人間がきめるべきかどうかは... 判断が難しい。

Q: 片寄 (LIST) 感性レベルでのメディア変換を目指しているのか?

A: いや、なるべく感性には踏み込まないようにしている。

C: それで正解だと今は思う。

Q: 後藤 (早大) メロディの抽出に用いたパラメータはどう設定したのか? 重み設定は?

- A: パラメータ数は6つ。選択結果が外れた要因を調べて、フィードバックさせる。80曲を対象にパラメータを設定したが、すべてを正しく選択するには設定できない。ある曲で選択できるようにパラメータを変えると、今まで正しかった曲で選択できなくなることがある。
- C: 長嶋 (ASL) 抽出したメロディのマッチングを取るのには簡単か? このデータベースを JASRAC に用意すれば、登録時のチェックに使えるのでは? アレンジに関わりなく、メロディの盗作チェックができそうに思われるが、
- (3) UNIX 上の MIDI シーケンサーとそのインタフェース
高田敏弘 (NTT)
記録: 増井 (富士通研)
- Q: 片寄 (LIST) イベントの管理にはタイムスタンプを使っているのか?
- A: MIsend というライブラリ関数があり、メッセージは一度カーネルにためられ、時間がくれば RS-232C ポートから送出される。MIDI 程度の精度なら最小限度まで十分管理できる。
- Q: カーネルに一度に送っても OK か?
- A: そうだ。メッセージを送る場合には、何秒間かデータ送出が行なわなければ、ユーザー側のプロセスをサスペンドさせるようにも設定できる。
- Q: 竹内 (NTT) どのくらいの精度なのか? RS-232C だから 1m sec 近くか?
- C: 長嶋 (ASL) MIDI では 320 μ sec に 1 つ、1 m sec で最大 3 発 (3 バイト)。
- Q: 竹内 RS-232C ポートでそこまでの精度が出るのか?
- C: 長嶋 SC-55mkII の場合は、(受けとったデータを) 内部で保留して発音を遅らせ、同時発音の精度を上げているらしい。
- Q: 後藤 (早大) MIsend の呼び出しは、以前のイベントのタイムスタンプが来るまでブロックされるとあるか?
- A: 設定で変更可能。たとえば 5 秒間はブロックしないように設定できる。
- Q: ブロックされる場合は、ノートオンで和音が届くと同時に鳴らないのでは?
- A: 完全にブロックすると駄目。ブロックされるまでの時間的余裕を設定することで、カーネル上のバッファ溢れを調節できる。
- Q: Mltimcreate, MIsend についてもそういう機能があるのか?
- A: 用意されている。
- Q: 森 (浜松ポリテク) Internet 上での広域 DB で Mosaic を利用する場合、Indigo を持っていないと受け手は MIDI データの再生ができないのか?
- A: 現状ではそうだが、テンボがヨクたっていいなら、Sparc でも MIDI 再生ができる。IBM-PC, Mac でもクライアントがあれば可能だろう。
- Q: midifile の解析ライブラリ: midifile 1.11 は SGI についているのか?
- A: SGI 添付でなく PDS である。MIDI 関係の各アーカイブサイトにも置かれている。
- Q: どのワークステーションでも利用可能なのか?
- A: Sparc 版の場合は、midifile ライブラリと、自前のコマンド送出部を用い、タイミングの制御はループで行なっている。
- (4) Conferences, Concerts, and Cocktail Parties: Besides Immersion
Michael Cohen (会津大)
記録: 平賀 (図情大)
- Q: 平賀 (図情大) 音の再生はヘッドホン、スピーカーのどちらによるのか?
- A: Circumfering ヘッドホンによる。したがって [発表時のような] スピーカー用には optimize されていない。
- Q: どのようなパラメータを操作しているのか。特に距離との関係は?
- A: いろいろなパラメータがあるが、考慮していないものも多い。Distance attenuation は、基本的には gain attenuation だけで処理している。したがってドライで残響のない環境を作っていることになる。
- Q: Lyon (慶大) このような技術を実現するための band width はいつ頃実用になるか?
- A: ISDN は次世紀には本格化するとすると、あと 6 年ということになる。ついでに、音源について clustering のようなことも考えられる。画像では pixel 集合が 1 つの単体を表したり、逆に pixel もさらに細かく分割できるのと同様、音でも音やチャネルの集合を 1 つの単位として扱うということだ。
- Q: 小坂 (NTT) copy and paste でクローンを数多く作ったときの聞こえを論じたが、検証実験は行なったか?
- A: 特にやってはいない。Precedence effect (Haas effect) の実現方法には、他にもいろいろある。Precedence effect 自体のシミュレーション、平均化処理、source $\rightarrow$ sink の手動割り当てなど。音楽は多重度が高いので、こういった研究に向いている。
- (5) Sinusoidal model の特性分析と音合成への適用
小坂直敏 (NTT)
記録: 平賀 (図情大)
- Q: 伊藤 (NTT) デモのサンプリング周波数は?
- A: 楽音は 48kHz サンプリングだが、情報は 44.1kHz 分である。分析の窓は 2048 に固定。音声は 12kHz サンプリングだが、情報があるのは 5kHz ぐらいまで。
- Q: 楽音は歪みが耳についた。サンプリングレートの利害が原因か?
- A: そうだろう。
- Q: 歪み量は?
- A: まだ現実の信号については測定していない。
- Q: MQ アルゴリズムの原論文とは位相項の近似式が違うのか?
- A: やったのは結局論文通り。日本ではまだ試されていないようなのでやってみた。
- C: 長嶋 (ASL) 音声、自然音については OK だが、楽器音は頭が歪んでいる。Sine 波 200Hz で 3Hz の偏差というのは、「楽器」のピッチ設定精度としてはおもしろ程度で誤差が大きく、せめて 9-10 bit の周波数設定精度がほしい。
- A: 実用面ではその通りだろう。とりあえず今の範囲では検証が目的である。
- (6) Time Varying Extensions to the Moore Phase Vocoder
Eric Lyon (慶大)
記録: 平賀 (図情大)
- Q: 小坂 (NTT) 音楽家の意見は聞いてみたか?
- A: 全般にインパクトはあるようだ。特に non-musician の反応が大きい。
- Q: 音楽家への普及は考えているか?
- A: 考えている。現在の 5 通りの processor を 20 通りぐらいに増やし、公開するつもりである。使い方は各人によって変わるだろうから、選択幅は大きいほうがいいだろう。