

人間の認知的特性に基づく音楽設計理論の構築に向けて

竹中 毅, 相澤 祐一, 鈴木晋太郎

音楽の設計は、他の人工物の設計と比べて、環境条件や目的を明示化することが極めて困難である。そのため、音楽の設計理論に対する科学的、工学的なアプローチは少なく、作曲行為は人間の暗黙知に大きく依存している。一方で、既存の音楽には時代や地域を超えた共通性が見られることから、音楽的秩序には人間の認知的・身体的特性に基づく必然性があると思われる。筆者らは、人間の認知的特性に基づいて、創発的に楽曲断片を生成することで、音楽の発生学的な根拠を理解することを目的としてきた。本稿では、筆者らの研究例を紹介するとともに、新たな音楽設計理論の可能性について議論したい。

キーワード：創発的設計, 認知的特性, マルチエージェント学習, 引き込み

1. はじめに

人間は、有史以来、音律や音階、リズムや和声など様々な音楽的要素を発明し、それらの組合せによって音楽的断片を創出してきた。したがって、作曲とは有限の音楽的要素の組合せによって、人間にとっての認知的・身体的な価値を最大化する解（音楽）を探索する最適解探索問題の側面を持っている。例えば、Simon[1]は、「音楽は、人工物の最古のものの一つ」とし、「作曲を一つのデザイン問題とみるならば、……、他のデザイン問題において行っているのと同じ仕事、……、に直面するのである」と述べ、作曲家と工学者が同じ創造的な活動であるデザイン過程を共有していると主張している。確かに、音楽は人間の感覚を通してのみ意味を持つ人工物であり、認知や感情といった人間的な価値と人工物の機能や構造との関係性を明らかにする上で、極めて有効なベンチマーク問題のひとつであるといえるだろう。

しかしながら、音楽の設計論なるものが、これまで

科学的、工学的に確立されてきたかという点、残念ながら理想とは程遠く、現在においても、作曲行為は人間の暗黙的な知識や行為に大きく依存している。その難しさを（工学的に）表現するならば、機械や建造物といった他の人工物の設計のように、物理的、客観的な環境条件を明確にできないことや、聴取者にとっての価値を定義するための目的関数がいくつも考えられ、極めて複雑な多目的最適解探索問題になっていることが挙げられるだろう。

ところで、現在の西洋音楽において、最も支配的な音楽理論は機能と声法と呼ばれるものであり、その原型は今から約300年前に確立された[2]。機能と声法によると、すべての和音はトニック、ドミナント、サブドミナントと呼ばれる3つの機能に分類でき、これらの結合によって楽曲が構成される。この理論は、音楽構造の一般性を理解する上で非常に魅力的であり、1930年代には、音楽学者であるシェンカーの原構造分析に影響を与え[3]、1980年代には、言語学者のチョムスキーの生成文法理論に啓発された言語学、人工知能、音楽学の専門家たちの協働によって、音楽構造分析のための理論（GTTM）が確立された[4]。しかし、これらは、基本的に、既存の楽曲の分析を目指すものであり、実際に作曲を行うには、他の多くの問題を明らかにしなければならない。つまり、音楽を理解すること（アナリシス）と創ること（シンセシス）の間には、大きな非対称性が存在するのである。

それでは、作曲という音楽の設計問題に対し、科学技術はどのように貢献できるのであろうか？ この問

たけなか たけし
産業技術総合研究所 サービス工学研究センター 大規模データモデリング研究チーム
〒135-0064 江東区青海 2-41-6
あいざわ ゆういち
東京大学 大学院工学系研究科精密機械工学専攻
〒277-8568 柏市柏の葉 5-1-5
すずき しんたろう
㈱電通 営業部
〒105-7001 港区東新橋 1-8-1

題に立ち向かうには、音楽学のみならず、心理学や脳科学、計算機科学や統計科学といった様々な分野を統合した学融合的なアプローチが不可欠となる。例えるならば、オペレーションズ・リサーチ研究が実践してきたように、既存のシステムを分析するための社会科学、そこに関わる人間的特性を理解するための人間科学、望まれる解を導出するための工学を統合した新たな学問分野の創出に向けた挑戦が必要となる。

これまで、情報学や計算機科学の音楽への応用として、遺伝的アルゴリズムによって解探索するモジュールを組んだ作曲支援システム[5]や、既存曲のデータベースからニューラルネットワークによって組合せ最適を行う自動生成システム[6]が提案されてきた。しかしながら、これらのシステムでは、人間であるユーザが評価を行ったり、教師信号としての既存曲断片との類似性を利用するものであったりするため、そもそも、多くの既存曲に見られるような秩序やルール（例えば旋律や和声進行）が、どのような人間的必然性によって発現したのかを理解することは難しい。

一方で、世界の民族音楽に目を向けると、地域や時代を超えて、音階や旋律、リズムなど、ある種の共通した秩序が観察される。つまり、そこには、人間的特徴に基づく何らかの必然性があると思われる。このような背景から、筆者らは、人間の認知的な特性を評価関数として導入し、創発的な計算論を利用することで、音楽の発生的側面の理解や新たな音楽の設計理論の構築を目指してきた。

続く2節では、音楽への創発論的なアプローチについて述べる。第3節では、研究例としてマルチエージェント学習を用いた和音列の生成に関する研究と物理モデルを用いた旋律の生成に関する研究を紹介したい。

2. 創発現象としての音楽

音楽は楽音から構成されているが、音楽を聴いているとき、個々の音を独立に認識することは少なく、フレーズやカデンツといった、ある程度の音集合から音楽的秩序を感じている。このような認知的特性は、ゲシュタルト性と呼ばれており、音楽心理学によって古くから指摘されてきた。

そこで、音楽を複雑系の問題として捉えてみると、要素となる音とそれらの時間的・空間的な配置によって実現される楽曲の間には、要素（楽音）と全体（楽曲）の循環関係による創発的図式[7]を想定することができる（図1）。また、配置された楽音はそれぞれ

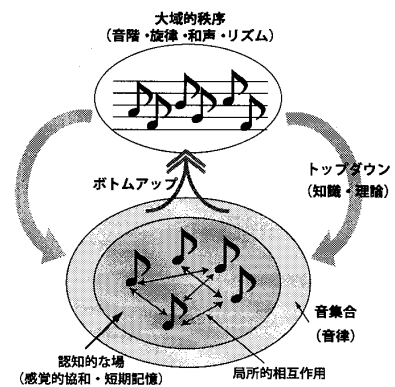


図1 音楽の創発的図式

が独立ではなく、人間の身体的・認知的特性に基づく相互依存的関係にある。したがって、音楽を創発現象と捉え、シンセシスの視点から理解することは、従来の楽曲分析や作曲活動の心理学的理解というアナリシスの視点において困難であった暗黙的領域の問題、すなわち楽曲に含まれる音楽的秩序や構造が人間のどのような認知的・身体的特性に基づいて創発しうるかという問題を扱うことが可能となる。

例えば、ある作曲者が先の機能と声法を用いて、トップダウンに音楽を創作しようとしても、和声進行以外の多くの部分に関して、試行錯誤を通じたボトムアップな過程が必ず存在する。ただし、その過程で作曲者が行っている価値判断は、極めて暗黙的であり、これまで定式化がなされてこなかった。筆者らの目的は、このような認知的な場における評価関数の存在を明らかにすることであるといえる。

3. 研究紹介

本節では、筆者らがこれまでに行ってきた2つの研究について紹介をする。1つ目は、音をエージェントとし、マルチエージェント学習を用いることで、使用音階や調性といった音楽的秩序が創発的に生成されるシステムを構築したものである。2つ目は、旋律が持つ様々なアクセントの周期的特徴に着目し、引き込み理論を用いて旋律が創発的に生成されるシステムを構築したものである。そこでは、聴取実験による生成曲の評価を合わせて紹介する。

3.1 マルチエージェント学習を用いた和音列の創発

我々が普段聴きなれている西洋調性音楽では、12個の利用可能な音集合（音律、ここでは平均律）のうち、ある7音で構成される長音階を基本として楽曲が

創作される。例えば、和音は、そのうちの3音（トライアド）で構成される。一方、日本のわらべうたなどの民俗音楽では、異なる音律から様々な5つの音が選択され、旋律やハーモニーが創られてきた。それでは、利用可能な音集合（音律）の中から、人間はどのように使用音を選び、音楽を創造してきたのだろうか？

そのような問題意識から、筆者らは、楽音をエージェントとして、エージェント間の相互作用の結果、音楽（和音列）を得るシステムを構築した。その際、人間の基本的な認知的特性として、感覚的協和（2音間の周波数比によって生じるうねりがもたらす心理的な協和の感覚）[8]と短期記憶、音の近接性（時間的に隣接する音の高さが近いこと）という計算可能な3つの評価関数を仮説的に導入した[9]。

図2に示すように、学習器を有する楽音エージェントは、横軸が時間、縦軸が音高である2次元空間において、離散的な時間ステップごとに他の2エージェントとの位置関係を観測し、どれだけ移動するかを行動として出力する。各エージェントの行動の結果、生成された配置を評価関数に基づいて評価したものを得点化し、報酬として与えることで、エージェントは、それらの報酬を最大化するような、和音列を生成する行動を獲得する。

その際、同一時間ステップにおける2音間の「空間的協和」に加え、一定時間の過去にわたる音と現在の音の間に生じる不協和に時間経過の割引率を乗じて定量化した「時間的協和」という新たな評価関数を設計した。これは人間の短期記憶を想定したものである。

まず、一般的な12平均律の環境において、計算機実験を行ったところ、空間的協和と時間的協和という2つの評価関数の導入によって、図3に示すように、12個の利用可能な音集合（平均律）の中から、選択的に5個、あるいは7個の音を利用することが学習された。さらに興味深いことに、インドで用いられる22音律（1オクターブに22個の音が存在する）を用

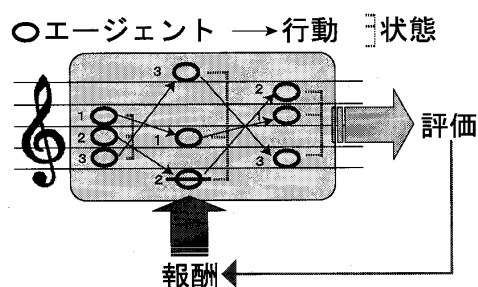


図2 マルチエージェントシステムの概要

いた環境では、その中から、5音を用いたインド音楽に実在する音階が選択的に利用されることがわかった。

これらの結果は、感覚的協和と短期記憶という人間の特性を背景として、西洋音楽における調性的秩序や民俗音楽における音階が生み出されてきた可能性を示唆するものである。また、音の近接性を組み合わせることによって、より旋律らしい音楽が創出されることも明らかとなった。本稿では詳細は述べられないが、これらの結果は、聴取実験によっても確認された。

さらに、音楽家に計算機実験と同一条件下での楽曲制作を依頼し、計算機実験の生成曲と比較することで、人間と提案モデルの相違点を検証した。作曲家による

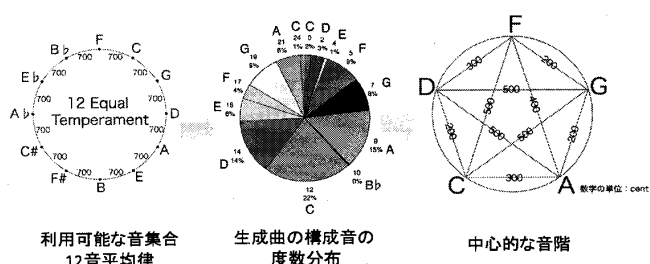


図3 生成曲の構成音の分布



図4 生成曲の一例（開始和音と最終和音を固定）

		1音目																						
		-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11
音高	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

		1音目																						
		-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11
音高	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

図5 計算機実験による生成曲(上)と人間による回答曲(下)の音の二次遷移

楽曲は、計算機による生成曲と比べて、はるかに魅力的なものであったが、例えば、生成曲と回答曲の個々の音の二次遷移に着目したところ、生成曲よりも回答曲では「上昇—下降」「下降—上昇」という遷移が避けられており、連続的でなめらかな旋律を実現していることが分かった。また、音楽家による作曲は、長音階の構成音を比較的均等に用いている傾向も見られた。この結果は、人間が、より高次の音楽的文脈を考慮していることを意味している。

音楽的文脈については、これまで音楽心理学や音楽学の分野において様々な議論がなされてきた。例えば、旋律の予期に関するナムアの暗意—実現モデルでは、人間は次の旋律に対する暗意（期待）と実現（あるいは裏切り）の連続によって、次の旋律を連続して予期しており、それが音楽的文脈の認知の基礎にあると考えた[3]。そこでは連続する3音の窓とクローザーにより、旋律の原型を8つに分類する汎様式的モデルを提案しており、これまでに音楽分析のための工学的な応用も試みられてきた。しかしながら、それらは音楽分析が目的であり、そのような文脈がどのようにボトムアップに創発されるかという問題はいまだ明らかにされていない。

続く3.2節では、旋律の時間的文脈に関する研究について紹介する。

3.2 引き込み理論に基づいた旋律の設計

旋律は、音高と音長という2つの離散的な量を持つ音符から構成されており、同様のパターンが繰り返し表れるといった特徴が観察される（図6）。すなわち、このような特徴は、時間と音高が相互依存的に設計されることで実現されているといえる。

そこで、パターンを生成する手法として、人間のタッピングによるリズム生成のメカニズムとして着目されている引き込み（Entrainment）に着目した[10]。引き込みとは、互いに固有振動数が異なる非線形振動

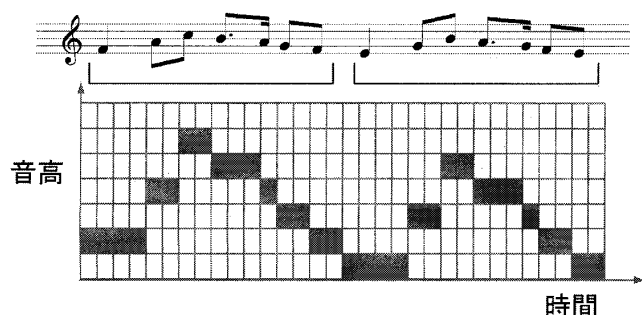


図6 旋律中のパターン

子が弱く結合している系において、相互作用の結果、振動子が同期する現象のことである[11]。このような系では、結合の強さや接続性によって、同位相、逆位相の引き込み、あるいは $m:n$ の振動数比に安定する整数比引き込みなど、様々なパターンが生成される。自然界では、大多数のホタルの明滅が同期するといった現象がその例である。

近年、音楽心理学においても、人間が旋律を聴取するときに、時間と音高を統合的に認知している可能性が指摘されており、そのメカニズムを説明するモデルとして引き込みが着目されている。例えば、Jonesらは、時間的側面としてのリズムのアクセントと音高に関するアクセントが統合される（Joint Accent Structure）ことで、旋律の記憶や認知に影響を与えていると説明しており、時間と音高の処理を説明するモデルとして Pitch/Time Entrainment model を提案している[12]。アクセントとは、周囲よりも目立って知覚される特徴点と定義され[13]、特に、発音（NO）や音高遷移方向の変化（CP）が代表的なアクセントである。

提案モデルでは、発音のタイミングを決める1個のNO振動子と音高遷移方向の変化のタイミングを決め

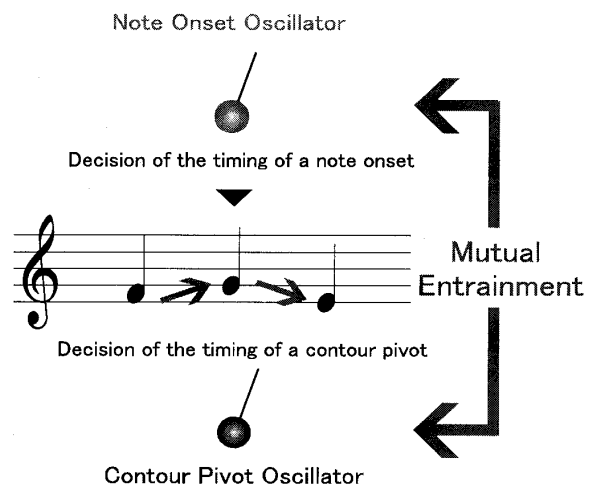


図7 振動子の引き込みによる旋律生成モデル

表1 生成条件

	NO 振動子	CP 振動子	結合
Case I	○	○	×
Case II	×	○	×
Case III	○	×	×
Case IV	×	×	×
Case V (提案手法)	○	○	○



図8 生成曲の譜面

る1個のCP振動子が、位相差によって相互作用をする。相互作用の結果、どのタイミングで発音するべきか、どのタイミングで音高遷移方向を変化させるべきか、といった時系列情報が生成され、隣接する音符の音高と近接する範囲内でランダムに音高が選ばれる。ここで旋律生成実験のためにNO振動子、CP振動子の導入・非導入、結合の有無によって表1のように条件を設定した。2つの振動子を導入し、かつ結合のあるCase Vが提案手法である。結果、例えば、図8のような楽曲が得られた。

設計手法の評価を行うために、得られた旋律と既存曲(Case VI)を刺激として、聴取実験を行った。聴取実験では、質問紙に5段階評定法に従って、最もあてはまると思われる尺度に○をつけさせた。各刺激の評点を全参加者間で形容詞ごとに平均したあと、同一条件の刺激間で平均化した平均得点を算出したところ、提案手法は、振動子の一方が存在しない条件に対して、まとまりや安定性について高い評価を得ていた。そこで、因子分析を行って主要な因子を抽出し、各刺激に対する印象をマップにした。因子分析の結果、「規則性のある・まとまりのある」といった規則性を表す因子と「速い・明るい」といった活動性を表す因子が抽出された。

図9に示すように、提案手法は、Case Iに対して、活動性に関する軸で高い評価を得た。また、音楽経験者に絞ると、そこに有意差が存在することも確認された。一方、生成曲に比べて既存曲は、両軸で高い評価を得ていることが分かる。用いた既存曲が調性音楽で

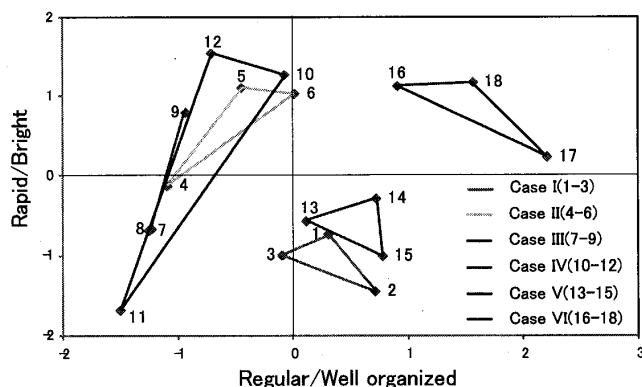


図9 因子分析の結果

あることを踏まえると、実験参加者は協和感や音高の絶対値について注意をしている可能性が十分に考えられる。そのため、今後は3.1節で紹介した強化学習などを導入することで、音高に関して解探索を効率的に行うようなシステムへと拡張したいと考えている。

ここで紹介した、旋律の音長と音高の同時的設計問題は、音楽の問題に限らず、広く、時間と空間の同時的設計問題と共通する。例えば、生産システムやサプライチェーンなどにおける、日程や工程の同時的最適化問題との類似点を考えるのは飛躍しすぎであろうか？ しかしながら、そのような異分野のアナロジーによって、新たな方法論が生まれることが、学融合的な研究にとって重要であると筆者らは考えている。

4. まとめと展望

本稿では、音楽の設計の問題について、創発論的な視点から、音楽的秩序の発生学的な理解と、認知的特性に基づいた新たな設計手法の確立に向けた試みを紹介した。既存の音楽理論は、限られたジャンル(例えば西洋調性音楽)の優れた楽曲の分析を通して、そこに見られる共通性を記号的に抽象化してきた。音楽理論は、主に音楽教育に用いられてきたが、そこでは、ある種の制約条件(禁則と呼ばれる)を示すことにとどまっており、実際に作曲を行う際には、その多くが人間の暗黙的な創造過程に依存することになる。しかしながら、音楽に携わるすべての人が抱える創造性への苦悩の原因を理解し、それらを緩和する意味においても、科学的手法を用いた音楽への構成論的なアプローチが必要ではないかと筆者らは考えている。

筆者らが目指す新たな音楽の研究方法論には、図10に示すように、これまでの既存の学問領域と違うプロセスが必要となる。すなわち、人間の認知的・身

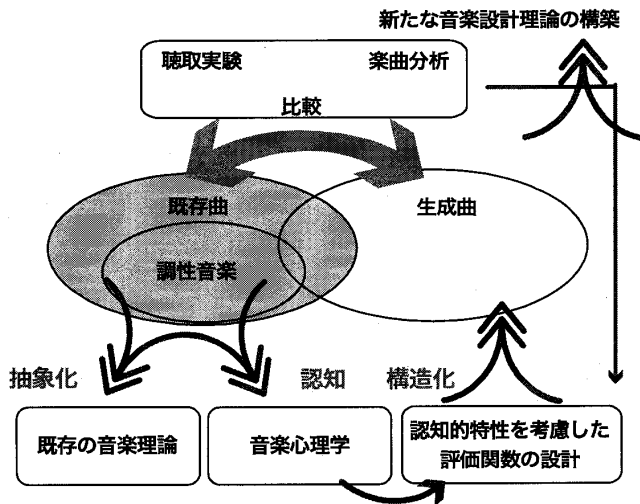


図 10 音楽の構成的理解に向けた研究方法論

体的特性に着目し、望まれる音楽的秩序を発現する基礎となる人間的特性に関する仮説形成を行い、創発論的計算などの技術を用いて実際に楽曲を構造化する。さらに、構造化された生成曲を評価し、仮説の検証を行うために、楽曲分析や既存曲との比較、聴取実験が必要となる。このような過程を経ることによって、これまでの手法では困難であった音楽の暗黙的な領域の構成的理解と、より人間的な音楽設計理論の構築につながると考えている。

また、このような問題意識と研究方法論は音楽の研究に限定されるものではないと思われる。例えば、経営工学においても、新たなシステムの開発には、既存のシステムの分析にとどまらない挑戦が必要となる。また、問題を解決するための方法論は無数にあったとしても、構築されたシステムは、実際に社会に投入されて、初めてその価値を発現する。すなわち、社会的受容を考慮した設計を行うには、人間的な価値の多様性と相互作用を通じた社会的価値の発現メカニズムを理解することが重要であると思われる。

筆者らは、現在、サービス工学という新たな分野において、生活者を中心とした、サービスの構成的理解に向けた研究を行っている。音楽についてもそうであるように、サービスの受け手は一律ではなく、個々人が多様な価値観を持っている。一方、近年の急速なグローバル化とネットワーク化により、複雑な社会的相互作用が起り、サービスの社会的な価値を見極めることが大変難しくなってきた。例えば、心理学は伝統的に人間や社会の普遍的な特徴を研究対象としてきた

が、情報化技術の進展によってモノやサービスが溢れる現在、価値の多様性の問題がますます重要になってきている。

そのような問題を考える上で、“人類最古”の人工物の一つである音楽の価値を考えてみることは、様々な人工物の設計問題に示唆を与えてくれるのではないかと期待している。

参考文献

- [1] H. A. Simon, “The Sciences of the artificial,” MIT Press, 1999. 稲葉元吉, 吉原英樹訳, 『システムの科学』, パーソナルメディア, 1999.
- [2] O. Allan, 「和声の歴史」, 白水社, 2004.
- [3] 波多野諄余夫編, 「音楽と認知」, 東京大学出版会, 1987.
- [4] F. Lerdahl and R. Jackendoff, “A generative theory of tonal music,” MIT Press, 1983.
- [5] A. R. Burton and T. Vladimirova, “Generation of musical sequences with genetic techniques,” Computer Music Journal, Vol. 23, No. 4, pp. 59-73, 1999.
- [6] J. A. Franklin, “Recurrent neural networks for music computation,” INFORMS Journal on Computing, Vol. 18, No. 3, pp. 321-338, 2006.
- [7] 上田完次編著, 「創発とマルチエージェントシステム」, 培風館, 2007.
- [8] W. A. Sethares, “Local Consonance and the relationship between timbre and scale,” Journal of the Acoustical Society in America, Vol. 94, No. 3, pp. 1218-1228, 1993.
- [9] S. Suzuki, T. Takenaka and K. Kanji, “Co-creative composition using Multi-agent Learning: Toward the emergence of musical structure,” Intelligent Autonomous Systems 9 (IAS-9), pp. 479-486, 2006.
- [10] 相澤祐一, 布袋田由理子, 竹中毅, 上田完次, 相互引き込みモデルに着目した音楽の創発的設計, 第9回計測自動制御学会 SI 部門講演会, 2008.
- [11] 蔵本由紀, 「非線形・非平衡現象の数理 リズム現象の世界」, 東京大学出版会, 2005.
- [12] M. R. Jones, H. M. Johnston and J. Puente, “Effects of auditory pattern structure on anticipatory and reactive attending,” Cognitive Psychology, Vol. 53, No. 1, pp. 59-96, 2006.
- [13] G. W. Cooper and L. B. Meyer, “The rhythmic structure of music,” The University of Chicago Press, 1960.