

【研究ノート】

音楽を理解するコンピュータ

—人工感性の実現を目指して—

山口直彦

Computer Recognize Music - For Establish Artificial Emotion -

Naohiko Yamaguchi

Abstract : Tonal Pitch Space(TPS) is a kind of theory of computational musicology based on classical theory of harmonics. However, current popular music is primarily composed by the jazz theory. Hence, the development of a new computational musicology based on the jazz theory is indispensable. As a challenge to realize artificial emotions, this article outlines the computational musicology, especially TPS and, furthermore, reports on some current researches on adapting TPS to the jazz theory.

Keywords : Computational musicology, Music informatics, Tonal Pitch Space, Theory of harmonics, Jazz theory, Popular music

1. はじめに

ディープラーニング（深層学習）という革新的技術が誕生し、コンピュータチェスの Deep Blue やコンピュータ囲碁の AlphaGo が大きな成果を残した事、そしてシンギュラリティ（Singularity: 技術的特異点）という言葉が世界をにぎわせた事で、人工知能、あるいは AI(Artificial Intelligence) という言葉はすっかり世の中に浸透した。ロボットの世界でも、人間と同じ空間で動作する協働型ロボットの開発が盛んになっており、多様な環境に柔軟に対応した制御を行うために AI は欠かせないものになっている。

しかし現在の人工知能にはまだまだ足りない部分がある。人とロボットが協働する中で、ロボットを単なる機械ではなく、本当の意味で人間の「仲間」として受け止めるために必要な事、それはロボットに「心」を感じ、お互いを「理解しあえている」という感覚を得られる事である。特に介護・医療・教育の分野でロボットが活躍するためには、ロボットを心から「信頼」するために必須の要素と言えよう。そのためには、ロボットが人間の「心」や「感性」を理解すること、すなわち「人工感性（Artificial Emotion）」の技術が求められるようになる。

本稿では「人工感性」の実現に向けた取り組みとして、コンピュータに音楽を理解させる「計算論的音楽理論」という試みについて概説し、報告者が現在検討している最新のA

アイデアについて解説する。

2. 人工知能と感性

2.1 これからのロボットには「感性」が必要になる

「ロボット」という言葉は、チェコの作家カレル・チャペックが1920年に発表した戯曲『R.U.R (Rossumovi univerzální roboti: ロッサム万能ロボット会社)¹⁾』で生まれ、その語源はチェコ語の「robota (強制労働)」とされる。それ以来ロボットは基本的に「人間の代わりに労働を行うもの」として捉えられてきた。実際、現代においてロボットが大きく活躍しているのは製造現場(工場)や物流(倉庫)などであり、大型のロボットがパワフルに、高速に動作して、人間がまねできない速度・精度で製品を大量生産したり、人間が持ち運べないような大型の荷物を運んだりしている。その代わり、これら産業用ロボットは安全柵に囲われ、人間を柵の外に追い出して安全を確保している。

その一方で、ソニーのaibo、ソフトバンクのPepper、iRobotのルンバのように、人間の生活空間で人間と共に活動するロボットも増えている。産業用ロボットの世界でも、安全柵を必要とせず、人間と共に動作する「協働型ロボット」が注目を集めている。

人間と同じ空間でロボットが動作する場合、ロボットが動作効率だけを追求した動作をすれば接触事故の危険性が高まる。従って人間に危害を加えないように考慮した動作制御が必要であることは当然である事に加え、人間が「安心して隣にいられる」ような振る舞いが求められる。例えば、人間にぶつからない「完璧な寸止め制御」が実現していたとしても、ボクサーの打撃練習のようにブンブンと猛スピードで動くアームがあれば、隣にいる人間は本能的に恐怖心を覚え、安心して作業ができないだろう。ここで大切なのは実際に人間へ危害を加えるか否かではなく、「人間に恐怖心を与えず、このロボットは(人間に危害を加えないと)信用できる」振る舞いである。「恐怖心」という人間の心情を汲んだ動作制御、言い換えれば「人間に危害を加えない優しさを表現する動作」が求められるのだ。それだけに限らず、用途によっては「安全確実な動作をしつつもキビキビとした動き」や「ゆったりと美しい所作」や「丁寧でおもてなしの心を持った動き」などが求められる事もあるだろう。

そのためには、ロボットに人間の「感性」を持たせる、すなわち「人工感性(Artificial Emotion)」が必要になる。現在の人工知能は人間のように「柔軟に物事を考える」働きを目指して進化しているが、「人工感性」は人間のように「感情を理解し、汲み取る」働きを実現させることを目指す。

そもそも、人間の感性や感情というものが身体内でどのように現れ、コントロールされているのかは、現在の脳科学をもってしてもまだ十分には解明されていない。「人工感性」を実現させようとする試みは、いわばコンピュータ内に感性のデジタルツイン(仮想空間で現実空間を再現すること)を作ることによって「人間の感性」の理解を相対的に深めようとする行為でもある。

2.2 AI 美芸研アジェンダ

人間の感性が機能する場はたくさんある。中でも芸術は「特定の材料・様式などによって美を追求・表現しようとする人間の活動。および、その所産。[3]」と説明されるように、人間に固有の活動と考えられてきた。

美術家で人工知能美学芸術研究会(AI 美芸研)に所属する中ザワヒデキ及び草刈ミカは、従来人間しか成しえないと考えられていた領域まで人工知能が進出しつつある現状を踏まえ、機械と人間による芸術(行為)と美学を表1に示すように4つの領域に分け、第IV領域(機械の美学で作られた機械による芸術)はあり得るのかという問いを人工知能美学芸術研究会の根幹となる問題意識(AI 美芸研アジェンダ)として示している[1, 2]。

もっとも、コンピュータ(電子計算機)と芸術の関係性は今に始まった話ではなく、電子計算機が開発されたかなり初期の頃から、何らかの形でコンピュータが関わる芸術活動というものは模索されている。本稿の主題である音楽に関して言えば、シンセサイザーやシーケンサによる音楽制作・演奏は、表1の第III領域(人間の美学で作られた機械による芸術)に相当するし、アルゴリズムミックコンポジション(一定の手続き(アルゴリズム)により作曲される音楽作品)は表1の第II領域(機械の美学で作られた人間による芸術)の一種とも考えられる。ただしこの場合、アルゴリズムの設計や選定を行っているのは人間自身のため、本当の意味で第II領域に該当すると言えるかどうかは議論の余地がある²⁾。

表1 AI 美芸研アジェンダ [1, 2]

	人間美学 Human Aesthetics	機械美学 Machine Aesthetics
人間芸術 Human Art	I	II
機械芸術 Human Art	III	IV

2.3 音楽情報科学と計算論的音楽理論

2.3.1 音楽情報科学

では、「機械美学」による音楽はどうすれば作る事ができるのだろうか。

コンピュータの知見を通じて「音楽」を取り扱う研究分野を総合して音楽情報科学または音楽情報学(Music Informatics)と呼ぶ。音楽情報科学の歴史は長く、世界初のコンピュータによる自動作曲作品「イリアック組曲」が発表されたのは1957年である。世界初の実用的なプログラム内蔵方式電子計算機であるEDSACが稼働したのが1949年、世界初の汎用電子デジタルコンピュータUNIVAC Iの発売が1951年であることを考えると、電子計算機黎明期からコンピュータを用いた音楽制作の取り組みが行われていた事がわかるだろう。

その後、シンセサイザの性能向上などと連動しながら、コンピュータは主に音楽制作

の道具として進化を遂げた。特に電子楽器の演奏情報を装置の間で転送・共有するための共通規格としてMIDI (Musical Instrument Digital Interface) が開発されたことで、カラオケや携帯電話の着信メロディなど、活躍の場を広げていった。90年代後半から2000年代以降は、インターネットの普及や iPod を筆頭とした大容量の携帯型デジタル音楽プレイヤー・スマートフォンが普及したことにより、個人が大量の音楽を所有・聴取することが当たり前になった。そのような状況から、音楽情報検索 (Music Information Retrieval : MIR) や音楽推薦の研究が盛んになった。また2007年にクリプトン・フューチャー・メディア社から発売された歌声合成ソフトウェア「初音ミク」(ただし歌声合成技術の基盤はヤマハ株式会社が開発した「VOCALOID2」による) [4] が記録的な売り上げを出したことで、歌声合成などの音楽情報科学の研究が大きく広がり、現在に至っている。

後藤真孝によれば、「音楽情報学が学問として目指すものは主に「音楽の理解・鑑賞」「音楽の創作・演奏」「音楽の流通・検索・推薦」「音楽による人間同士のコミュニケーション」「音楽に内在する性質」等を工学(役に立つモノを作る)あるいは科学(真理を明らかにする)の立場から明らかにすることである」と説明する。そして具体的な課題として「音楽に関する信号処理、採譜、音源分離、識別、分析、理解、検索、推薦、分類、流通、同期、変換、加工、要約、作曲、編曲、作詞、演奏、伴奏、楽譜認識、楽音合成、歌声合成、生成、支援、符号化、視覚化、インタラクション、ユーザインタフェース、データベース、アニメーション、ソーシャルタグ等の様々な課題が関係する。」と述べている [5]。

2.3.2 計算論的音楽理論

音楽について考える際、「そもそも音楽とは何か」という問いは極めて基礎的な疑問に見えて、実際には非常に答えることが難しい問題である。音楽は(物理現象としての)音を媒体とした活動であり、複数の音を時間軸及び周波数軸上に分布させたものであるが、それらが全て音楽として知覚されるわけではない。どのように音を選び、並べれば「音楽」という創作物として認識されるのかを示す知識体系を、一般に「音楽理論」という。ただし絶対的に正しい音楽理論というものが存在するわけではなく、その内容は文化背景や時代などによって異なる。現在「音楽理論」といった場合は、ヨーロッパを中心に構築・発展してきた音楽(いわゆる西洋音楽)に由来する理論(楽典・和声学・対位法など)を指すことが一般的である。

西洋音楽理論の基本は、いわゆるクラシック音楽を作曲するために体系立てられたクラシック音楽理論である。クラシック音楽理論もまた時代と共に変化し、特に近現代では古典的なクラシック音楽理論からかなり逸脱したのも音楽として受け入れられている。また1990年代初頭から、西洋音楽がアメリカに伝わって黒人文化と交わりジャズとして発展する中で独自の音楽理論を形成し、ジャズ音楽理論(ポピュラー音楽理論とも)として現代の商用音楽の世界で広く使われている。特にバークリー音楽院において体系的にまとめられた「バークリーメソッド」と呼ばれる理論体系が有名である。

コンピュータによって作曲を行ったり、楽曲の分析を行う場合には、前提知識として「音楽理論」をプログラム上で表現しておく必要がある。しかしこれらの「音楽理論」は、人間が学習して使用する事を前提に構築されてきたものであり、また論理的に構築された理論というよりも、経験則の集合体である。従って定義があいまいであったり、解釈が一意

に定まらなかったり、矛盾する内容が含まれていたりすることもある。そのため、この「(人間用の) 音楽理論」を直接プログラムに翻訳してコンピュータに理解させようとする、問題が生じてしまう。そこで、コンピュータが理解できるような数学的なモデルとして構築し直された音楽理論が必要となる。これを「計算論的音楽理論」という。

代表的な計算論的音楽理論として、以下の3つがある。

暗意 - 実現モデル (Implication-Realization Model) [6, 7]

旋律 (メロディ) を分析対象とし、音高の遷移パターンを3音から成る基本類型に分類する手法。Eugene Narmour が提唱。

生成音楽理論 (GTTM : Generative Theory of Tonal Music) [8]

旋律 (メロディ) を分析対象とし、旋律を木構造に簡約・集約することで上位構造を分析する手法。Fred Lerdahl 及び Ray Jackendoff が提唱。

Tonal Pitch Space (TPS) [9]

和声 (コード) 進行を分析対象とし、和音間の「進行しやすさ・進行の自然さ」に対応した「和音間距離」を定式化する手法。GTTM を補完する理論として Fred Lerdahl が提唱。

本稿ではこのうち、Tonal Pitch Space (以下 TPS) に着目して議論を進める。

3. TPS を拡張する試み

先に示した3つの計算論的音楽理論を含め、従来検討されてきた計算論的音楽理論のほとんどはクラシック音楽理論をベースに構築されていて、ジャズ音楽理論に基づいた理論は少ない。

しかしながら、現在世界中で消費されている音楽の多くはジャズ音楽理論に基づいて作曲されている。従って、より実用的な計算論的音楽理論を構築するにあたっては、ジャズ音楽理論を取り上げることが不可欠であると筆者は考える。そこで本研究は「ジャズ音楽理論」を計算機上に実装することを試みる。ジャズ音楽理論はコード (和声) 進行を中心に考える理論であるため、同じく和音に着目している TPS が大きな参考となる。TPS をさらに拡張する形でジャズ音楽理論に照応する計算論的音楽理論を構築する事を目標とする。

3.1 TPS とは

前提として、TPS の考え方を先に概説しておく。

TPS の中心となる概念は「和音間距離 $\delta(x \rightarrow y)$ 」である。与えられた2つの和音 x, y が $x \rightarrow y$ の順番で鳴らされた (和声進行) した際、その進行の自然さに相当する指標を「和声間距離」という値で定式化する。これにより、和声学でカデンツ (終止形) と呼ばれるパターンを基礎として理論づけられてきた和声進行が、数式によって表されることとなる。

和音間距離を式 (1) で表す。なおここで x, y はそれぞれ与えられた和音、 $\delta(x \rightarrow y)$ は和声進行 $x \rightarrow y$ の和音間距離を表す。(表記は Lerdahl の原著 [9] ではなく、先行研究 [10, 11] に基づいている)。この値が小さければ小さいほど、 $x \rightarrow y$ の進行は自然であるとされる。(この「自然」はあくまでも指標であって、音楽的に許されるコード進行か否かを絶対的

に決定できるものではない。表現上の効果を狙って、あえて和音間距離の大きいコード進行を使用する場合もあるし、逆に和音間距離が小さいコード進行ばかりでは、平坦でつまらない曲となる可能性もある。)

$$\delta(x \rightarrow y) = \text{region}(x, y) + \text{chord}(x, y) + \text{basicspace}(x, y) \quad (1)$$

式(1)で示された和音間距離は、 $\text{region}()$ 、 $\text{chord}()$ 、 $\text{basicspace}()$ という3つのサブ関数の和となっている。

3.1.1 region 関数

region 関数は、和音 x, y がそれぞれ置かれている調を比較し、図1に示すような調の5度圏 [15]³⁾ 上でどれだけ離れた位置に存在するかを region 距離として返す関数である。

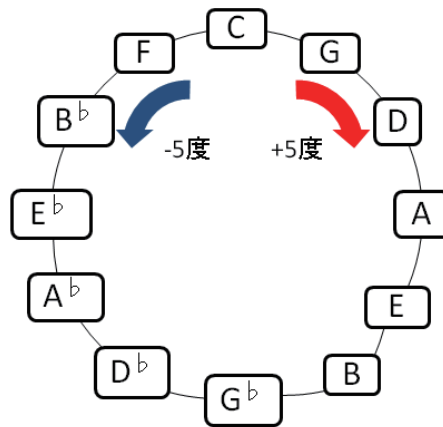


図1 調の5度圏

調の5度圏はC～Bまでの各調の主音が、隣同士完全5度の関係になるよう、円環状に配置したものである。図1では時計回りに1ステップ進むと完全5度上昇、反時計回りに1ステップ進むと完全5度下降するようになっている。これにより、ある調(例えばG)の右隣は属調(Gに対するD)、左隣は下属調(Gに対するC)を表す。

region 関数は、この調の5度圏上で、 x, y の置かれている調が何ステップ離れているかを返す。なお、図1に書かれた調は全て長調であり、短調の場合は平行調⁴⁾の関係を用いて長調に変換する必要がある。

なお、この方法では与えられた和音の置かれた調が遠隔調の関係にある場合、求められる和声間距離が小さくなりすぎるため、さらに調性空間と呼ばれる空間を考慮した算出を行う手法も提唱されている。

3.1.2 chord 関数

chord 関数は、和音 x, y が図2に示す和音の5度圏⁵⁾上で、どれだけ離れているかを chord 距離として返す関数である。

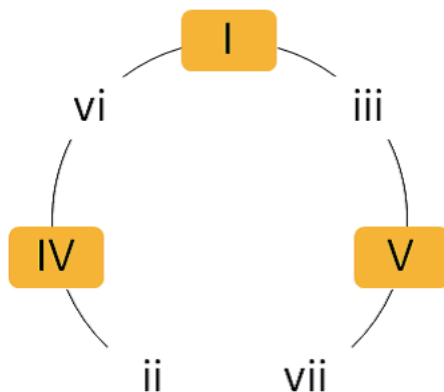


図2 和音の5度圏

和音の5度圏は調を構成する音それぞれを根音とする三和音（ダイアトニックコード）を円環状に配置したものである。図2は調の主音をIとした音度で表記している。

ダイアトニックコードの中でもI（主和音、トニック）、IV（下屬和音、サブドミナント）、V（属和音、ドミナント）は和声進行の根幹を成す重要な和音であり、主要三和音と呼ばれる。和音の5度圏は、主要三和音の間を、副三和音（ダイアトニックコードのうち、主要三和音でないもの）を、隣同士が代理和音の関係となるように並べられている。

和音 x, y が同じ調である場合は、単純に和音の5度圏上の距離を求めれば良いが、両者の置かれている調が異なる場合は、一方を他方の調へ転調した上で距離を比較する必要がある。またその際、距離が求められない場合があり、その場合は十分に大きな値を返すこととする（すなわち、異常な和声進行とみなす）。

3.1.3 basicspace 関数

TPS では、TPS は1つの和音を12次元のベクトルで表現し、これをベーシックスペースと呼ぶ。ベーシックスペースの各要素はC～Bまでの半音階にそれぞれ対応し、Level.a～Level.eの重みを持っている。ベーシックスペースは調名と根音の音度のペアで一意に表現でき、以下根音の音度と太字の英語表記による調名（長調は大文字、短調は小文字）をスラッシュではさんで表記する。図3に、I/C（コードネームではCmaj）を表したベーシックスペースを示す。（原著[9]ではピッチクラス（CからBまでの各半音に割り当てた0～11の数字）を用いてベーシックスペースを図示しているが、読みやすさのため本稿では音名を用いて表記している）。

Level.a	C										
Level.b	C						G				
Level.c	C		E				G				
Level.d	C	D	E	F		G	A	B			
Level.e	C	Db	D	Eb	E	F	Gb	G	Ab	A	Bb

図3 I/Cのベーシックスペース

3.2 TPS をジャズ音楽理論に適用するために

先述の通り、TPS はコード進行を分析対象とした理論であることから、同じくコード進行を中心に考えるジャズ音楽理論と親和性が高い事が考えられる。しかし実際には、TPS を直接ジャズ音楽理論と結びつけて使用するには、以下の問題点が存在する。

非調構成音の扱い あるレベルとなっている音は、それよりも下位のレベルも含まなくてはならない

- すなわち、「和音構成音は調構成音でなくてはならない」ため、ダイアトニックコードや V_7 といった限定的なコードにしか対応できない。
- ジャズ音楽では、和音は一般に 4 和音以上で鳴らされる事が多いが、常用される I_7/C すら、調構成音でない音 (B \flat) を含むため、記述できない (ベーシックスペース上では、図 5 内、記号 ϕ で示した部分に「抜け」が生じる。)

和音構成音の重みづけ TPS では、主音、5 度音以外の和音構成音は全て Level.c とされている。

- すなわち、7th などの付加音が、3 度音と同等の重みづけがされている
- 3 度音は和音の響きを決める重要な音であり、7th などの付加音は機能上 Omit 可能であり、これを等価値として良いのか疑問である。

chord 関数の改訂 上記に指摘した「和音構成音の重みづけ」の問題に関連し、chord 関数もまたダイアトニックコードにしか対応していない。

Level.a	C									
Level.b	C				G					
Level.c	C		E		G			B \flat		
Level.d	C	D	E	F	G	A	ϕ	B		
Level.e	C	D \flat	D	E \flat	E	F	G \flat	G	A \flat	A
									B \flat	B

図 5 I_7/C のベーシックスペース

表 3 TPS-ExJ におけるベーシックスペースの各レベル

レベル	解説
Level.a	和音構成音のうち、根音が含まれる
Level.b	和音構成音のうち、根音及び 5 度音が含まれる
Level.c	和音構成音のうち、根音・5 度音・3 度音が含まれる
Level.d	和音構成音が含まれる
Level.e	調を構成する音が含まれる
Level.f	C~B まで全ての半音階が含まれる

距離を算出する。その結果を表4に示す。

計算によって得られた距離が最小となったダイアトニックコード（表4で太字で示した部分）と、与えられたコードとを連結し、図7のように和音の5度圏を拡張する。

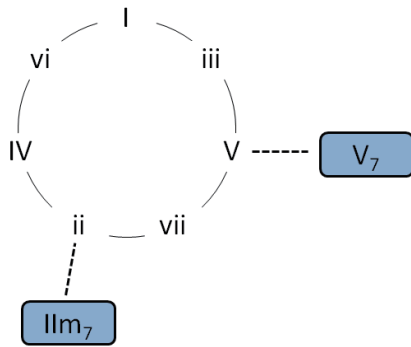


図7 拡張された和音の5度圏

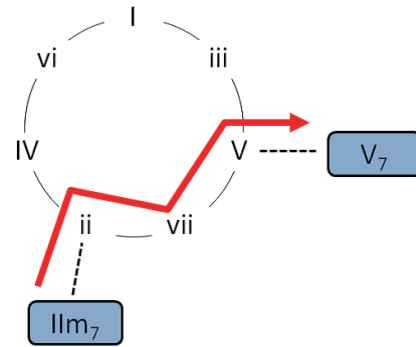


図8 拡張された和音の5度圏での距離算出

拡張された和音の5度圏上で、与えられた和音間の距離を求め、 $\text{chord}(\text{II}m_7/C, \text{V}7/C) = 4$ と出力する（図8）

3.3.3 TPS-ExJの評価と問題点

提案手法をプログラムによって実装して基礎的な実験を行った結果、当初の目的通りTPSが扱える和音の種類は増え、また和音間距離の算出も行う事ができた。これによりTPS-ExJの目的はある程度達する事ができた。その一方で、以下のような問題点がまだ残っている。

分解能の不足 探索空間（表現できる和音の範囲）が広がった分、和音間距離の分解能が相対的に悪くなってしまい、適切な和音進行解釈を行うことが困難になってしまった。特に転調に対するコストが低すぎ、人間の解釈では通常転調と捉えないような進行に対しても、転調を行っているとして解釈する場面が見受けられた。

スケールの導入 ジャズ音楽理論では調とコードの間に入る概念として「スケール」が導入されたり、あるいは（クラシック音楽理論に由来する）調の代わりに「モード」と呼ばれる概念が導入される場合がある（「スケール」「モード」とも、音階（C～Bまでの12半音）の中から、楽曲を構成する音を抽出したものと捉えることができる）。TPS-ExJはこれを適切に表現することができない。TPSの問題点「非調構成音の扱い」への対処として、調構成音でない音が和音構成音となった場合に、ベーシックスペース上一時的にその音を「調構成音」とみなす処置は、この「スケール」「モード」に対する暫定的な対処であるが、十分な対応とは言えない。

心理学的評価の不足 そもそも、TPS及びTPS-ExJが算出する「和音間距離」が、人間の感性が評価する「和音の解釈・評価」とどのように関連するののかという点について、心理学的な検証がまだ不十分である。

4. ジャズ音楽理論に照応する計算論的音楽理論の構築に向けて

ここまでの成果を踏まえ、TPS-ExJ にさらに改良を加えることを検討している。また、TPS-ExJ の実用性について、心理学に基づく実験・評価を行う事を検討している。

4.1 TPS-ExJ の改良 (案)

TPS-ExJ をより実用的にするため、「和音間距離のベクトル化」「任意スケールへの対応」「転回形への対応」の3点について改良を検討している。

4.1.1 和音間距離のベクトル化

前節にて示したTPS-ExJ の問題点として「分解能の不足」を挙げていた。これはそもそも、TPS-ExJ の出力を、和音間距離という1つのスカラー値に集約しようとしているために起こっている事であると考えられる。

昨今の人工知能技術(ニューラルネットワークなど)の状況と照らし合わせて考えれば、和音間距離というスカラー値に集約するのではなく、和音の解釈に関係する個々のパラメータを集めたベクトル値、「和音関連ベクトル」としてとらえた方が、自然な解釈となるのではないだろうか。(ただしベクトルの各成分は、値が大きくなるほど和音進行が「不自然である」と解釈する方向に定義することとする)そして、和音間距離を求める際には、和音関連ベクトルと各成分の重みベクトルを内積した値を得るようにする。このようにしておけば、重みベクトルを調整することでTPS-ExJ の評価を柔軟に調整することができ、機械学習アルゴリズムによって重みベクトルを学習させることも可能になる。

4.1.2 任意スケールへの対応

音のピッチは原則として基本波の周波数で決まり、任意のピッチ(周波数)の音を作る事ができる。西洋音楽理論では楽曲に使用する音として1オクターブ(周波数2倍)の範囲を12等分⁶⁾に離散化して使用する。離散化した12個の音から楽曲中で主に利用する音を、ルールに沿って選び出す。西洋音楽理論においては「調」という概念がこれに相当し、12個の音を主音(中心音)とした長調/短調の2種類が定義されるので全部で24種類の調が存在する。

TPSにおいては、長調/短調の概念がregion()関数やbasicspace()関数に導入されているが、これ以外の音階(スケール)を表現できない。例えばジャズやブルースでは「ブルー・ノート・スケール」と呼ばれるスケール(図9)が利用される。ブルー・ノート・スケールは同主音の長調を基準とし、第3音・第5音・第7音を半音下げた音(ブルー・ノート)を併用するスケールとされている。TPSにおいては、このブルー・ノートが調外音となってしまうため、ルール上記述することができない。TPS-ExJにおいては、和音の中に調外音が存在することを許容しているため記述自体は可能であるが、和音の解釈を行うにあたり、妥当な判断が行えているかは疑問が残る。ブルー・ノート・スケール構成音は同主音の短調(ハ調自然短音階、Cマイナー・スケール)とも構成音が似ており、長調/短調という概念だけではスケールを十分に解釈することができない。

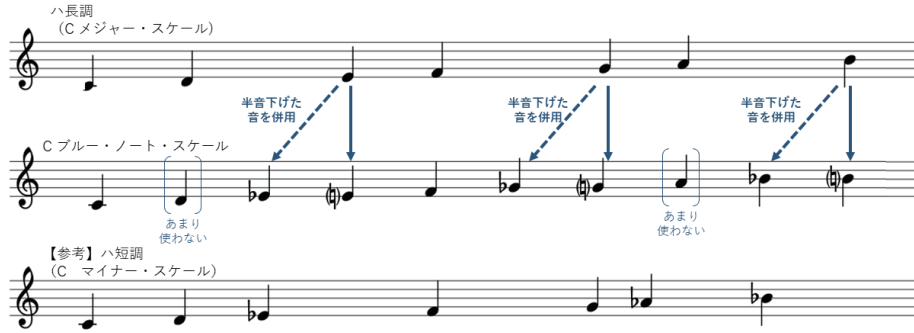


図9 ブルー・ノート・スケール

具体的な対応方法としては、ベーシックスペースの Level.e (調を構成する音が含まれる) を調ではなくスケールに置き換えてしまう方法が考えられる。しかし調とスケールを単純に置き換え可能な、等価なものとして良いかどうかは検証の必要がある。または Level.e と Level.d (和音構成音が含まれる) の間にもう一つ、スケール構成音のレベルを挿入する方法も考えられるが、これは逆に長調/短調という概念と両立しえないスケールも存在するため、スケールと調の関係をどのように定義するか検証する必要がある。

またどちらの方法を採用しても、旧来の「調」という考え方に基づいて作られている region 関数や basicspace 関数を、「スケール」という概念にどう対応させて関数の出力 (評価値) を得るかが課題となる。

4.1.3 「転回形への対応」

一般に和音は、基本となる音 (根音) を一番低い音にして、その上に第3音 (メジャーコードの場合は長三度 (根音の4半音上)、マイナーコードの場合は短三度 (根音の3半音上))、さらにその上に第5音 (7半音上) … という形をとる (このように根音の上に音を積み重ねる形を和音の基本形という)。これに対し、和音を構成する音を組み替えて、根音以外の音が最低音になるようにしたものを「和音の転回形」という (図10)。



図10 コードと転回形の例 (Dmaj 及び D⁹)

音楽理論上、和音の転回は和音の持つ機能を変えないとされている。しかし演奏上は和音の構成音が同じでも、音を重ねる順番や高さを変えることで聴感上の印象は大きく異なるため、目的に応じて音の積み方を演奏者が調整したり、作曲者が指定する事もある。ジャ

ズ音楽理論ではこれをコード・ヴォイシングという。例えばコード構成音が複数コードに渡って順次上昇（下降）するようにヴォイシングする「クリシェ」という手法（図11）はよく利用される。

図11 クリシェの例

現在の TPS-ExJ ではコード・ヴォイシングを記述することも、ヴォイシングを考慮した分析を行うことができない。和音の構成音だけでなくその積み方も記述・分析できるように改良することで、より実用的な楽曲分析が可能になると考える。

具体的な対応方法として、TPS-ExJ に新しくヴォイシングスペース (VoicingSpace) という概念と、ヴォイシングスペースの距離を求める voicingspace 関数を提案したい。

ヴォイシングスペース (VoicingSpace) として、12次元のベクトルを用意し、ベースックスペースと同様、各要素を C ~ B までの半音階にそれぞれ対応させる。和音構成音のうち一番低い音（コードネームで最低音が指示されていない時はルート音）に3、二番目に低い音（同第3音）に2、三番目に低い音（同第7音）に1の値を入れ、他は0とする。I/C = Cmaj のヴォイシングスペースを図12に示す。この場合ヴォイシングに対して明確な指示が行われていないため、ルート音であるCを最低音とし、第3音、第7音であるE,Gがその上に積まれるものとする。

Cmaj	3	2	1									
	C	Db	D	Eb	E	F	Gb	G	Ab	A	Bb	B

図12 Cmaj/C のヴォイシングスペース

また、V/ConB = Gmaj/B のヴォイシングスペースを図13に示す。この場合はベース音がBと指定されているため、最低音がBとなるように転回している。

Gmaj/B		2				1				3		
	C	Db	D	Eb	E	F	Gb	G	Ab	A	Bb	B

図13 Gmaj/B のヴォイシングスペース

voicingspace 関数は、与えられた2つの和音からの変化量を求める関数である。ヴォイシングスペースに示された3つの音それぞれの移動量を求め、重みとかけ算したものを合計して voicingspace 関数の出力とする。例えば voicingspace (Cmaj/C → Gmaj/B) であれば、図 14 のように、移動量は最低音からそれぞれ 1,2,0 となるので、重みと積和して

$$\text{voicingspace}(\text{Cmaj/C} \rightarrow \text{Gmaj/B}) = 1 \cdot 3 + 2 \cdot 2 + 0 \cdot 1 = 7 \quad (3)$$

を得る。

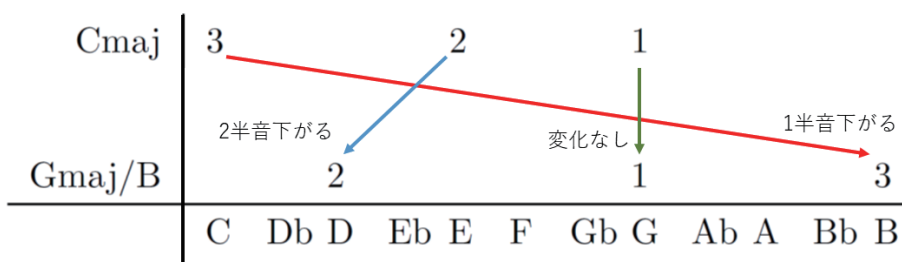


図 14 voicingspace 関数の計算例

ほとんどの和音は3つ以上の音で構成されているため、ヴォイシングスペースの扱いは上記の定義で対応できる。ただしロックンロールなどで用いられるパワーコード（3和音から3度音を抜いたもの、omit3とも表記する）については構成音が2つだけになるため、パワーコードにも対応させる場合はどのように処理すれば良いか、検討を要する。

4.2 TPS-ExJ の心理学的実験・評価

4.2.1 音楽の評価と計算美学

TPS は和音間距離 $\delta(x \rightarrow y)$ が小さいほど、コード進行 $x \rightarrow y$ の進行が自然であるという前提であるが、先述の通り和音間距離を最小にすれば良い音楽になるわけではない（仮にそうだとすれば、最初から最後まで和音を変えず、ワンコードで進む曲が最良ということになってしまう）。

音楽の複雑さと快感情に関する関係については、Berlyne (1971) が最適複雑性モデル (optimal complexity model) という考え方を示した。これは複雑さを横軸、快感情 (好感度) を縦軸に曲線をプロットすると、山なりの逆 U 字曲線を描く (すなわち、最大の快感情を得られる最適な複雑さが存在する) というモデルである [16, 17]。

計算美学を研究した川野洋は、作品を部分要素が n 個並んだ時系列と考え、各要素の持つ個別情報量 (サプライザル) $H_{(t)} (t=1, \dots, n)$ と、みる者の情報知覚容量を C としたとき、作品の美的価値 V は式 (4) で表されるとした [18]。(ここで f は $|H_{(t)} - C|$ に比例して大きくなる関数とする)

$$V = \sum_{t=1}^n f(|H_{(t)} - C|) \quad (4)$$

これはすなわち、各クロックにおける個別情報量 $H_{(t)}$ と情報知覚容量 C の差が小さいほど、 C という能力を持つみる者にとって作品の美的価値が大きくなる事を意味する。音楽

はまさに作品の要素を時系列に沿って提示する芸術であるから、川野の考え方は音楽にも適用できよう。情報知覚容量 C というある種の情報量の最適値が仮定されていることから、最適複雑性モデルとも考え方は矛盾しない。

ただし、川野の著書 [18] では個別情報量と情報知覚容量の差 $|H_{(t)} - C|$ を $t = 1 \dots n$ の範囲で、つまり作品の開始から現在までの情報を全て合計して作品全体の美的価値を求めている。これは同書で川野が主に絵画を対象として考察を行っており、絵画は全体を一度に見ることができるためにこのような式になったと思われる。音楽の場合は時系列ごとに過去の刺激は消失していくため、作品全体の価値を1つの値に定めるのではなく、人間の短期記憶に相当する時間分の情報のみを使って算出した美的価値が時間を追って変化していくと考えた方が自然であろう。実際、川野の別書籍 [18] では、ごく短いながらも音楽を対象とした考察が行われており、そこでは音楽をマルコフ・チェイン⁷⁾とみなしている。これを式 (4) に合わせて数式化すれば、短期記憶可能な時間を T クロックとして (すなわち T 次マルコフ過程として)

$$V(t) = \sum_{\tau=t-T}^t f(|H_{(\tau)} - C|) \quad (5)$$

のように考えるのが望ましいと考える。

4.2.2 TPS-ExJ の心理学的評価方法の検討

さて、TPS-ExJ が算出する和音間距離は、コード進行の自然さに相当する値であって、Berlyne の最適複雑性モデルにおける (音刺激の) 複雑さや、川野の美的価値モデルにおける個別情報量と相関性があると推定される。TPS-ExJ が算出する和音間距離が、人間が感じる音楽的な心地よさや美的価値と相関性を持った評価を出せているのか、またより良い相関を持たせるために TPS-ExJ をどう改良すべきかを検討するため、心理学的なアプローチでの実験・評価が必要となる。

具体的な手法としては、一対比較法 (より正確にはシェッフエの一対比較法 (浦の変法)) を使用する。

被験者に2つの試験和音列 P, Q を順番に提示し、どちらが好ましい (または自然な) 和音列であったか点数をつける (例えば5段階であれば2点・1点・0点・-1点・-2点)。ただし、試験和音列だけを鳴らしてしまうと、調性感覚がつかめないままの評価になってしまう可能性があるため、試験和音列のまえに調性を示すための基準和音列を提示する必

図 15 試験パターンの例

要があると思われる。図 15 は、ハ長調及びト長調で $vi - ii - V - I$ と $vi - IV - V - I$ という 2 つの試験和音列を比較する試験パターンの例である⁸⁾。試験和音列の前に、それぞれの調性を提示するための基準和音列 $I - IV - V - I$ を付加している。

実験にあたってはできるだけ多くの組み合わせで調査を行いたい。しかし調査対象の調性が K 個、それぞれで N 個の試験和音列を使用すると仮定すると、 $KN(N-1)$ 通りの試験パターンを作ることになる。被験者の負担を考慮し、1 試験パターンあたり回答に 1 分かかるとして、前後の説明や準備も含めて 1 時間程度で納めると仮定すると、調性 2 種類・試験和音列 5 個（試験パターン 40 通り）が現実的な量ではないかと考えられる。

このような対比較法によって得られた和声進行の評価と、TPS-ExJ 評価値を比較することで、TPS による評価値の妥当性が検討できると考える。

5. おわりに

本稿では「人工感性 (Artificial Emotion)」の成立を大きな目標とし、その実現に向けた取り組みとして、コンピュータに音楽を理解させる「計算論的音楽理論」という試みについて概説した。中でも特に Tonal Pitch Space (TPS) という理論に着目し、これをジャズ音楽理論に拡張した TPS-ExJ という理論の解説と、さらなる改良案について解説と検討を行った。

本稿で提示した TPS-ExJ の改良案については、今だ仮説の状態であり、実装の試みや実験・評価などの研究プロセスを進めて妥当性を評価していく必要がある。本稿の内容について、ご意見や有益な情報などあればぜひ筆者にお送りをご希望。また本稿を通じて、音楽情報科学というちょっと変わった研究分野があるという事を知って頂き、その面白さに触れて頂ければ幸いです。

謝辞

本研究は工学院大学大学院において、指導教官である管村昇教授の指導の下行われた筆者の修士論文が元になっています。管村先生には長期間にわたり、仔細に至るまで御指導・御鞭撻頂いたことを深く御礼申し上げます。

本研究を進めるにあたっては、プロベシストである山口康広氏に音楽理論について多くのご教示を頂きましたことを感謝いたします。

情報処理学会音楽情報科学研究会、日本音響学会音楽音響研究会の皆さまには、研究発表の折に有益な情報を多数いただきましたことを感謝いたします。

注

- 1) 「エル・ウー・エル」と読む
- 2) そもそも人工知能美学芸術研究会 (AI 美芸研) の問題提起は、中ザワが「人工知能が作ったアートと今いわれているものは、人工知能ではなくて人間が、人工知能という道具を使ってつくったものに過ぎないではないか！」という怒りを出発点として生まれたと述べている [1, 2]。

- 3) 一般には単に5度圏と呼ばれるが、後述の和音の5度圏と区別するため、このように呼ぶこととする
- 4) 同じ調号を用いる長調と短調の関係
- 5) 先述の調の5度圏と違い、隣り合う音同士の関係は5度ではないため、この図を5度圏と呼ぶ妥当性にいささか疑問が残るが、ここでは文献[10, 11]の表記に沿っている
- 6) 厳密にはこの表現は音律として現在主流の12音平均律を使用した場合のみ成り立つ。これ以外の音律を採用した場合、等分とは言えない
- 7) マルコフ・チェイン、またはマルコフ連鎖は、次ステップの状況が、現在及び有限過去の情報からのみ決定される確率過程（マルコフ過程）のうち、離散時間で表されるものの総称
- 8) この例では適宜和音の転回をして、和音間で音の跳躍が少なくなるようにしている

参考文献

- [1] 中ザワヒデキ、草刈ミカ「対談：「人工知能美学芸術展」をめぐる」、人工知能学会誌、33巻6号、pp.700-710, 2018年11月
- [2] 人工知能美学芸術研究会『人工知能美学芸術展』、人工知能美学芸術研究会、2019年12月
- [3] デジタル大辞泉「芸術」、<https://dictionary.goo.ne.jp/word/%E8%8A%B8%E8%A1%93/#jn-66566>, 2021年8月8日確認
- [4] 剣持秀紀、大下隼人「歌声合成システム VOCALOID - 現状と課題」、情報処理学会研究報告、Vol.2008-MUS-74, 2008年2月
- [5] 後藤真孝「新しい〇〇情報学：5. 音楽情報学」、情報処理、51巻6号、pp.661-668, 2010年6月
- [6] Eugene Narmour『The Analysis and Cognition of Basic Melodic Structures』, The university of Chicago press, 1990
- [7] Eugene Narmour『The Analysis and Cognition of Melodic Complexity』, The university of Chicago press, 1992
- [8] Fred Lerdahl, Ray Jackendoff『A Generative Theory of Tonal Music』, The Massachusetts Institute of Technology, 1983
- [9] Fred Lerdahl『Tonal Pitch Space』, Oxford University Press, 2001
- [10] 坂本鐘期、東条敏「Tonal Pitch Space を用いた楽曲の和声解析」、情報処理学会研究報告、Vol.2009-MUS-80, 2009年5月
- [11] 坂本鐘期「Tonal Pitch Space を用いた楽曲の和声解析」、Master's thesis, 北陸先端科学技術大学院大学、2010年3月
- [12] 山口直彦「非調構成音を含む和音への対応を目的とした TPS の拡張：ジャズ音楽理論への適用を目指して」全国大会講演論文集、2010年3月
- [13] 山口直彦「非調構成音を含む和音への対応を目的とした Tonal Pitch Space の拡張」、Master's thesis, 工学院大学大学院、2011年3月
- [14] 山口直彦「非調構成音を含む和音への対応を目的とした TPS (Tonal Pitch Space) の拡張：ジャズ音楽理論への適用を目指して」、情報処理学会研究報告、Vol.2011-MUS-89, 2011年4月
- [15] 小方厚『音律と音階の科学 新装版 ドレミ…はどのように生まれたか』、講談社ブルーバックス、2018年5月
- [16] 林原理恵、尾田政臣「和音進行の複雑さが快感に及ぼす影響」、映像情報メディア学会技術報告、2009年3月
- [17] 林美都子、ウィリアムズ信介「音楽経験者をもっと複雑な曲を好むのか—音程と好ましさに関する逆U字曲線を探る—」、2019年度日本認知科学会第36回大会、2019年9月

[18] 川野洋『コミュニケーションと芸術』、塙書房、1968年10月

[19] 川野洋『コンピュータと美学—人工知能の芸術を探る』、東京大学出版会、1984年12月

山口直彦 東京国際工科専門職大学 工科学部 情報工学科 助手