

新しい音楽スタイルの創出のための 音楽理論生成アルゴリズム

東京芸術大学大学院美術研究科先端芸術表現専攻

田中 翼

目次

第 1 章	序論	1
1.1	研究目的	1
1.2	自動作曲における創造の問題	1
1.3	本研究のアプローチ	2
1.4	本論文の概要	3
第 2 章	旋律の形式文法の生成	6
2.1	旋律スタイルと形式文法	6
2.2	関連研究	7
2.3	形式文法的アプローチ	9
2.4	多声音楽のための非同期モデル	11
2.5	旋律文法の生成手法	16
2.6	生成曲の評価実験	26
2.7	章のまとめ	29
第 3 章	感情を表現する旋法の生成	31
3.1	旋法と感情	31
3.2	関連研究	32
3.3	強化学習	33
3.4	提案手法	36
3.5	実験	40
3.6	第 2 のモデルについての考察	50
3.7	章のまとめ	53
第 4 章	音程スケールと群の生成系	55
4.1	旋法の変種としての音程スケール	55
4.2	音程スケールの定義	55
4.3	音高と音程の双対性	56
4.4	音程スケールの概念を適用できる楽曲例	58
4.5	\mathbb{Z}_n の代数的構造	60

ii 目次

4.6	音程スケールの選択方法	64
4.7	音列操作と音程スケール	67
4.8	作曲への応用	69
4.9	章のまとめ	70
第5章	サウンドファイルの対位法	72
5.1	音符からサウンドファイルへ	72
5.2	ミュージック・コンクレート	74
5.3	サウンドアーカイブ	74
5.4	音楽情報検索	75
5.5	コンピュータと音楽理論	75
5.6	サウンドファイルの対位法の構成	78
5.7	章のまとめ	83
第6章	結論	86
	参考文献	89
	付録	94

第 1 章

序論

1.1 研究目的

20 世紀以来、情報技術は芸術の領域に対して影響を及ぼしてきた。特に音楽は数として取り扱いやすいこともあり、芸術の諸分野の中でも科学技術との親和性が高く、電子音楽の創作、音響合成、自動作曲、音楽生成支援技術、音楽情報検索などの研究が行われてきた。現在では、音楽プログラミング言語 Max (Max/MSP) や歌声を生成するボーカロイドなどをはじめ、音楽生成の支援技術は実用化し、様々なソフトウェアを誰もが手軽に使用できる状況にある。また自動作曲の工学研究においては、様々なジャンルの楽曲スタイルをコンピュータで模倣して生成するような研究が多く行われてきた。だが、このような、人間の創作をコンピュータが支援したり、人間の考え出した音楽スタイルをコンピュータによって模倣するというような方向性の研究においては、芸術的なオリジナリティの源泉を人間に求めることになり、コンピュータや計算を用いる意義が利便性だけに限定され、矮小化されてしまいかねない。そこで本研究ではコンピュータの使用が、補助的な道具として、あるいは人間の代行としての利便性にとどまらず、いかに新しい音楽スタイルを生成し創造性に寄与できるかという問題意識を基本に据える。そして、個々の作品を深層から規定する音楽理論的な構造を対象とした数理的な生成手法を開発し、それによって新しい音楽スタイルの創出を試みる。

1.2 自動作曲における創造の問題

近年の情報技術の発展はビッグデータの時代と呼ばれる特筆すべき状況を生み出している。Web 上のテキスト情報や商業的な顧客情報等、様々な電子データが大量に蓄積されるようになり、それを活用する自動処理技術が研究・開発されてきた。音楽においても、MIDI 等の電子化された楽譜情報、演奏情報が存在し、それに対する確率統計的手法やデータマイニング手法を応用し、音楽スタイルや演奏スタイルを模倣するような研究が盛んである。しかし、このような研究の多くは、人間の作り出した一次情報に依存し、後からそれを利用する方法を考えていることになる。ここで私が問題だと考えるのは、このようなタイプの音楽の自動生成が既存の楽曲の情報に依存する点において二次創作的であるという点であり、一から新しい音楽を

2 第1章 序論

作り出すような創作とは質が異なるのではないかということである。

さらに、ここには自動作曲にとっての象徴的な問題が横たわっている。一般に、芸術や音楽の創作は創造的な仕事と見なされる傾向が強いといえるだろう。他方で自動化というのは、人間でなくても機械に代行させられるという意味で、創造性とはかけ離れた意味合いをもつ。では、その二つの語が結合した自動作曲というものは、創造的なものかそれともそうではないのかという疑問が生じる。

ここで、自動作曲における自動という概念を整理するため、自動化を二つの側面に分けてステレオタイプとして把握しておくことは有用である。一方は、既存の音楽理論のもとでの楽曲のリアライズや、既存の楽曲スタイルの模倣、芸術家が新たに考えた作曲手順の遂行作業を手計算ではなく自動的に行うといった、ルーティンワーク的作業をコンピュータが代行する消極的な自動化である。他方は、未だ存在しない新しい形式の楽曲をコンピュータに生成させ、芸術的な創作の根幹となる部分をコンピュータに行わせようとする積極的な自動化である。前者は主に工学的な自動作曲の研究において主流となる考え方であり、芸術作品としての新しさは要求せず、どこかで聞いたことのあるような違和感のない楽曲さえ自動生成できれば良いという価値観をベースとするものである。ここでは、創造というよりも既存のスタイルの模倣をいかに上手く行うかということが問題となる。他方、後者は芸術作品としての新しさこそをいかにコンピュータの計算力を活かして作り出すことができるかというチャレンジングな課題である。

両側面は完全に分けられるものではないとして、もどちらに重点を置くかによって自動作曲の概念の意味が変わってくる。本論文で私が目標とするのは、後者の創造的な側面に多少なりとも踏み込み、既存の楽曲と似たような作品の再生産で満足するのではなく、コンピュータを用いて初めて生み出されるような新しいスタイルの作品を生み出すための手がかりを作り出すことである。

1.3 本研究のアプローチ

新しいスタイルの楽曲を生み出すために本研究で私が着眼するのは、楽曲そのものよりも楽曲の背後に横たわる音楽理論である。従来、音楽理論はコンピュータが生成する対象として考えられることは少なく、作曲家や音楽理論家が頭で考えて作り出されてきたものである。そして自動作曲の研究においても、音楽理論は研究以前的前提としてすでに与えられているものとして扱わがちである。しかし、本研究では、音楽理論そのものをコンピュータによって生成する対象として扱う。その理由は、音楽理論を与えられたものとして考えている限り、それに基づいて生成される楽曲は既存の音楽スタイルの枠から外に出ることが困難であるからである。従来の工学的な研究がこうした新しい楽曲スタイルを生成する研究に踏み込みにくかったのは、工学系の研究者が、音楽理論そのものは自分の研究領域ではないため音楽理論家に任せて、自分は既存の理論を受け入れた上でどのように楽曲を生成するかというふうの問題を設定しがちであったためではないかと私は考えている。しかし私は、音楽理論を生成対象とすることによってこそ、新しいスタイルの作品を創造するための自動化の研究に道が開けるだろうと

考える。

本論文で扱う音楽理論は、具体的には、旋律法、旋法、音程スケール、サウンドファイルの対位法の四つであり（最後の二つは私自身が考案した概念である）、それらの抽象的な構造を生成の対象とすることで新しい楽曲スタイルを作り出そうとするのが本論文全体に通底する研究アプローチである。

だが、どのようにすれば新しい音楽理論的な構造を生成できるだろうか。生成するものは単に未知のものであれば良いわけではなく、無秩序なもの、無価値なもの、無根拠なものと同様に区別されるような音楽にとって本質的に重要なものでなければ意味がない。それをどのように創発あるいは選択できるかが問題となる。そこでの困難は、新しいものを作ろうとするためには、既存の実例を用意してそこから学習するような通常の方法論とは別の方法論が必要になることである。そこで私が案出したのは以下のような方法論である：

第2章においてとった方法論は、出来上がった楽曲の模倣ではなく、人間がオリジナルなスタイルを創造するプロセスこそを模倣するというものである。人間は、最初から優れたスタイルにたどり着くのではなく、試行錯誤を経て、失敗や発見をしながら、徐々に自身のスタイルを形成して行くものだろう。また、当人または他の誰かが既に獲得した型やミーム（文化的な遺伝子）を受け継ぎ、発展させるような歴史性も持つ。私は、そのようなプロセスこそをコンピュータ上で模擬することでオリジナルな創造が可能になるのではないかと考え、漸次的に学習する機械学習のモデルを構築した。第3章においては、機械学習のモデルに人間とのインタラクションを導入し、人間の感情を適応的にモデルに反映させる方法論をとった。第4章では、音楽理論のもつ数学的な構造に着眼して望ましい条件を導き出し、その条件を満たすような構造を、列挙、スクリーニングによって絞り込んでいく方法論を用いた。第5章では、多数の素材を用意し、その特徴を分析した上で、望ましい素材の組み合わせ方を特徴の検索によって見だし、その情報に基づいて楽曲を構築するという方法論をとった。

1.4 本論文の概要

以上のようなアプローチに基づく様々な方法論を用いて、本論文では新しい楽曲スタイルを作り出すための研究を行った。本論文の概要は以下の通りである：

第2章では、個々の旋律のスタイルを規定する旋律法を生成対象とした。そこでは、作曲家 Olivier Messiaen の、音型の引用と合成による作曲プロセスを参照し、それを形式文法の観点からとらえ直した。そして進化や淘汰を模擬した遺伝アルゴリズム、およびそれを利用したルールの生成手法であるクラシファイア・システムを用いることで、形式文法規則の創発を試みた。また、複数の旋律の同時共存する多声音楽にも対応するため、声部ごとの旋律法を同時共存させるモデルを構築をした。このモデルには、ルール数および声部間での旋律法の共有／非共有という二つのパラメータが組み込まれている。そのパラメータを変化させることで、生成された旋律法の評価がどのように変化するかを実験した。結果、ルール数を小さく抑えた場合に、旋律スタイルの独創性や統一性などの観点において評価が向上することや、声部間の複数の旋律法の共存が、旋律の豊かさを向上させるという効果が示された。

4 第1章 序論

第3章では、旋律法の基礎でもあり音楽理論の最も基本的な部分の1つである旋法 (mode) を生成対象とした。そこでは、旋法が感情表現と密接な結びつきを持つことに着目し、特定の感情を良く表現できるように人間からのフィードバックを用いて旋法を自動的に感情に適応させる強化学習モデルを構築した。この手法の利点は、適応的な方法を用いることで、通常の心理実験のように予め固定した無数の旋法を逐一評価する必要がなくなり、あらゆる種類の旋法の中から未知の旋法を発見できる可能性がある点である。強化学習は試行錯誤による成功や失敗の経験から学習していくプロセスをモデル化したものであるが、実際、特定の旋法が使用されるようになる歴史的経緯は、全ての可能性を比較して選ぶというのではなく、長期間にわたる漸次的な適応の過程だと考えられる。構築したモデルは、そのような過程を短時間でシミュレートすることを意図するものである。ただし、モデルのトレーニングを行う人間にはそれなりの負担がかかる。そこで、負担を軽減するための工夫として複数の関数近似モデルを構築して比較した。評価実験として、喜び、悲しみ、恐れ、優しさの4つの種類の感情に対する旋法の適応を行った結果、感情の表現力をもつ旋法を学習することができた。この実験結果においては、未知の旋法が生成されたことに加えて、喜びの旋法としてミクソリディア旋法が生成されたり、喜びと悲しみの旋法として、それぞれ陽旋法と陰旋法 (日本の伝統的な旋法) が埋め込まれた旋法が生成されるなど、既存の旋法の意義の発見や学習における文化的な影響を暗示するような興味深い結果を含むものであった。また、関数近似モデルの定式化から得られた副産物として、旋法の概念のヴァリエーションとしての「音程スケール (interval scale)」という概念が見いだされた。

第4章では、第3章で見いだされた音程スケールの概念のもつ可能性をより深く調べるため、楽曲分析と数学的な考察を行った。そこでは、音程スケールの概念の適用例と解釈できる楽曲として、György Ligeti の「ピアノのためのエチュード第二番」および Alban Berg の「ヴァイオリン協奏曲」を挙げて分析を行い、この概念が作曲に効果的に用いられることを例証した。そこからわかるのは、音程スケールは、通常の旋法の形式に基づかない無調音楽においても、旋法や調による色彩の差異のようなものを表現するために用いられる可能性があるということである。さらに私は、音程スケールが興味深い代数的構造をもつことに着目して数学的な考察を行い、この概念が無調性に結びつきやすいことを示す定理や、この概念を作曲に応用するために有用な数学的命題を見いだした。最後に、それらの考察結果として導かれた条件を用いて音程スケールの列挙、スクリーニングを行うことにより、望ましい音程スケールを生成し、作曲に応用した。

第5章では、電子音楽の時代に対応した音楽理論の必要性から私が案出した「サウンドファイルの対位法」という概念を提示し、生成対象として扱った。対位法は、複数の旋律を同時に組み合わせるための理論であるが、サウンドファイルの対位法とは、通常の音符を単位とした対位法をサウンドファイル単位素材として扱えるように拡張したものである。音符は楽器の出せる音の範囲で自由自在に書くことができるのに対し、サウンドファイルを用いた場合は、素材となるサウンドファイルしか用いることができない。使用する素材しだいで得られる音とその組み合わせの仕方も異なってくるため、素材に合わせた形で対位法を構成する必要がある。それを実現するための方法論として、サウンドファイルの音響分析によって素材として

のサウンドファイル断片の特徴づけをし、そこからの検索によってサウンドファイル断片の良い組み合わせを発見する枠組みを提案した。これにより、従来は創作者が主観的に音を組み合わせていたミュージック・コンクレートに対して、より客観的な音の組織化方法を提示した。また、音声データの量が飛躍的に増加している近年、それを有効活用するためには、いちいち音声を耳で聞いて確かめるという時間のかかる作業を自動化し、必要な音声ファイルだけを探索して見いだすという手法が不可欠になるだろう。その意味でも、そうした自動処理技術を電子音楽の作曲法と融合させる音楽理論としてのサウンドファイルの対位法は、今後ますます重要になると考えられる。

第2章

旋律の形式文法の生成

2.1 旋律スタイルと形式文法

旋律は音楽的要素として最も重要なものの一つである。本章では、新しい音楽スタイルの生成という目的のために、旋律のスタイルを規定する音楽理論としての旋律法に焦点をあて、その自動生成の手法を提案する。旋律スタイルを扱う研究においては、旋律スタイルを形式文法として定式化するアプローチの研究が多く行われてきている。したがって本研究でも文法的なアプローチを踏襲する。ただし、従来は既存の楽曲をうまく分析できるような形式文法を求める問題が扱われることが多かったが、本研究は旋律の形式文法を自動生成することで、新しいスタイルを作りだそうとする試みである。

旋律の文法を自動生成し、そこから本格的な音楽作品の生成を行おうとする際には、解決すべき主要な問題が二つある。第一の問題は多声音楽をいかに文法で扱うかという問題である。多声音楽は言語や単旋律のように単純な一つの線状のシーケンスではないため、それをどのように文法で表現するかは難しい問題である。第二の問題は、人間が試行錯誤で文法を設定するのでも、既存の作品からの学習によるのでもなく、未知のスタイルの文法をいかに生成するかという問題である。

これまでの文法的なアプローチの自動作曲研究においては、これらの問題はあまり重点的に研究されてこなかった。本章では、これら二つの問題への取り組みとして、多声音楽の一般的な文法モデルを構築し、その具体的な文法規則としての旋律法を自動生成することで新しい旋律スタイルを作り出す手法を提案する。

新たな旋律法を生成するためには既存曲の実例から統計学習するなどの方法をとるのは困難である。そこで本研究では、できあがった楽曲よりもむしろ、人間の作曲者がオリジナルなスタイルを創造するプロセスに着目する。人間は作曲において、最初から優れたスタイルを持っているのではなく、試行錯誤を経て、失敗や発見をしながら、徐々に自身のスタイルを形成して行く。また、当人または他の誰かが既に獲得した型やミーム（文化的な遺伝子）を受け継ぎ、発展させるような歴史性も持つ。そのようなプロセスこそをコンピュータ上で模擬することで、オリジナルな創造が可能になるのではないかと考えられる。

旋律法における、そのような型（ミーム）の受け継ぎをモデル化するにあたっては、作曲家

Messiaen の、音型の引用と合成による作曲プロセス [1] や、遺伝子を通した生物の進化を模擬した遺伝アルゴリズムが参考になると考えられる。本研究ではまず前者を応用しつつ形式文法的な旋律モデルを構築する。そして、そのモデルのもとで、新しい旋律スタイルの生成を音型の形成ルールの整合的な集合を発見する問題として定式化を行い、モデルの要素としての文法規則をクラシファイア・システムと呼ばれる遺伝アルゴリズムをベースとしたルール生成手法を応用して自動生成することで、新しい旋律法を創発させる手法を提案する。またそこでは、多声音楽にも対応するため、声部ごとに異なる（または同じ）旋律法を同時共存した形での生成を可能にするモデル化を行う。

ここで、本研究で取り扱う多声音楽を、ある声部間の音楽的制約を満たしながら、複数の独立した声部の旋律が同時進行する音楽として定義しておく。ただし各声部は同時に二音以上発音することはないものとする。また、声部間の音楽的制約は人間が手動で設定するものとし、自動生成の対象ではないことを予め断っておく。つまり対位法的（声部間的）な側面はモデル化のみで自動生成の対象とはせず、あくまで旋律的な側面に自動化の焦点を当てるということである。主に想定している楽器の編成は、人間の奏者または MIDI によるピアノであり、曲のデータは声部数分の、音高（音階上の度数）と音価の組を要素とするリストとして与えられる。このリストは適宜 MIDI や楽譜に変換される。

本章の構成は以下のとおりである。2.2 節では、本研究と既存研究との関係について述べる。2.3 節では、単旋律を対象とし、その生成過程を、書き換え規則に基づく成長プロセスとして捉える枠組みについて記述する。2.4 節では、第一の問題（多声音楽の文法的な扱い方）に対し、単旋律の生成モデルを、非同期的、並列的に拡張した多声音楽のモデルを提案する。2.5 節では、第二の問題（新しいスタイルの生成）に対し、クラシファイアシステムを用いて文法規則を獲得する手法を提案し、実際に生成した楽曲を紹介する。2.6 節では生成曲の評価実験について述べる。2.7 節では本章のまとめと今後の課題について述べる。

2.2 関連研究

音楽の中に文法を見いだそうとする研究は多数存在している。特に言語学者 Chomsky の生成文法は大きな影響力をもっており、音楽学者 Schenker の調性音楽の階層的な楽曲分析を生成文法と結びつけた GTTM [2] や、ジャズのコード進行のモデル化、民族音楽の旋律のモデル化等の応用がなされている [3]。

作曲に重点をおいた文法的なアプローチの研究としては、L-System を用いた旋律生成の研究が盛んに行われてきた [4, 6, 5, 7, 8]。L-System は、元々は生物の細胞分裂や植物の成長過程のモデル化に用いられた文法モデルであり、様々な目的対象を記号列としてコード化し、書き換えルールとして形式化された文法を記号列に反復適用することで、記号列を成長発展させていくシステムである。L-System は終端記号と非終端記号の区別のない一種の生成文法とも考えられ、決定論的なもの、確率的なもの、文脈依存的なもの、パラメータ付きのもの等、様々なバリエーションが考案され、音楽にも応用されている [3]。L-System 等の書き換え規則を用いた方法はシンプルさ、高速性、汎用性等の美点があり、音楽以外にも様々な物事の生

8 第2章 旋律の形式文法の生成

成モデルとして応用されている。特に近年ではコンピュータ・グラフィックスの需要の高まりから、自然景観や建物の外観等を高速に生成するプロシージャル技術として盛んに研究されている [9, 10]。多重音を扱った L-System の研究としては DuBois による [8] がある。DuBois は与えられた旋律に対して文法規則を適用することにより、リアルタイムに伴奏を導出するような枠組みを提案した。この枠組みでは、元の旋律に対して、規則による変形を施したものによって逐次的に応答するような伴奏部が生成される。しかしこの枠組みでは、元の旋律に伴奏部が従属するために伴奏部の独立性が弱かったり、伴奏部の声部数が変化するという点で、ポリフォニーよりもホモフォニーに近いと考えられ、本研究とは対象とする形式が異なる。また、本研究では必ずしもリアルタイム性を重視せず、作品の楽譜を最終的に生成できれば良いと考える。また、同じく多重音を扱った編曲の研究として、多重音を含む音楽オブジェクト間の包含関係や二項演算を定義し、編曲へ応用した [11, 12] がある。これらの研究では、ある楽曲およびその編曲例を参照した演算によって未知の楽曲に対する編曲を行う枠組みが提示されているが、作曲の問題への応用の可能性については言及されていない。

その他に、楽曲コーパスの中から文法を自動的に見つけ出す文法推論やスタイルを模倣の研究がある。それらを楽曲の生成につなげる研究の例として、ジャズの旋律に内在するマルコフ的な文法を楽譜から統計学習する [13] や、可変長の文脈を用いた文脈依存的文法により無用なルールの学習を最小限に抑え、原曲のスタイルへの忠実性を保とうとする研究 [14] 等が挙げられる。これらの研究は、旋律の文法の獲得を目的とする意味では本研究の立場に近いものである。ただし、これらは、学習対象の楽曲への依存度が強く、元の曲の断片がそのまま出現することが多い等、本研究の目的とするオリジナルな楽曲スタイルの生成には不向きだと考えられる。

文法的なアプローチの自動作曲の既存研究を見渡したとき、未だ十分に解決されていないと思われる重要な問題がある。それは多声音楽をどのように扱うかという問題である。生成文法におけるツリー状の階層構造や、L-system における一本の記号列や、マルコフ的なシーケンスといったものは、単旋律やコード進行を表現するには適しているが、独立した複数の系列が同時進行する多声音楽の構造を扱うことは困難である。多くの研究が単旋律を主な対象としていることも、そこに起因している面があると思われる。また多声部を書き換え規則で表現する例が示されている [5] 等の研究も存在するが、そこでは後に論じるように、複数の独立した状態系列を一まとめにして多声部を扱っている。このような方法をとると、旋律間の独立性が損なわれるか、状態の数が膨大化して文法が過度に複雑になるか、どちらかに陥る恐れがある。そのため、この問題を解決するには新しいモデルが必要だと考えられる。

文法的なアプローチ以外での多声音楽の生成研究としては、対位法のルールを満たす解を探索する研究 [15] や、確率モデルの統計学習によるもの [16] 等が存在するが、それらは古典的な音楽理論のもとで楽曲をリアライズする研究であり、オリジナルな音楽スタイルの生成を目的とする本研究とはスタンスが異なる。

本研究で文法規則の生成に用いるクラシファイアシステム (CS[17, 18]) は、if-then ルールの集合を遺伝アルゴリズム (GA) を用いて探索するシステムであり、視覚的な形態のデザインのための規則を自動生成する研究 [19] 等にも応用されている。書き換え規則は if-then ルー

ルとして扱うことができるため、CS は本研究にとって有用性が高いと考えられる。遺伝的手法を旋律の文法の生成に用いた既存研究に [7] がある。この研究では、突然変異および交配の演算が文法ルールの変形、多様化に用いられているのみであり、GA による文法ルールの探索は行われておらず、それについては今後の課題とされている。

以上をふまえ、本章で提示する研究の独自性は次の三点にある。(1) 文法的アプローチと多声音楽の生成との間のギャップを埋めるための一つのモデルを提示すること。(2) 旋律の文法の生成にクラシファイアシステムを応用すること。(3) 生成文法や L-system とは異なり、旋律生成への文法的アプローチを、人間の認知的な側面や、既存の楽曲の分析や、自己相似性の観点からではなく、作曲家の旋律作成プロセスの実例から根拠づけること。

2.3 形式文法的アプローチ

2.3.1 単旋律の既存モデル

書き換え規則は、生成文法や L-System 等の文法において用いられる生成規則であり、様々な物事の生成モデルの文法に用いられてきた。図 2.1-A (上) に示されるように、書き換え規則は predecessor と呼ばれる左辺の記号列を successor と呼ばれる右辺の記号列に置き換えるものである。この規則を、図 2.1-A (下) のように、最初に与えられた Axiom と呼ばれる記号列に対して適用することで、次世代の記号列を生成する。そして、その新しい記号列に対し、さらに同じ規則の適用するというように、記号列を段階的に成長させて行く。図 2.1-B は、生成された記号列の第三世代を左から順にグラフィカルに解釈したものである。この図形はヒルベルト曲線と呼ばれ、空間を効率的に充填する図形として数学者ヒルベルトが発案したものである。この解釈で、“F” は一定の長さ前進、“-” は左へ 90 度回転、“+” は右へ 90 度回転を表す。“X” および “Y” は解釈時には無視される。また、Prusinkiewicz は図 2.1-B のヒルベルト曲線を、左下をスタート地点として、X 軸の移動を発音、Y 軸の移動を音階上の移動と解釈し、図 2.1-C のような旋律を生成した [4]。このように、書き換え規則は旋律の生成のための文法としても用いることができる。

2.3.2 旋律文法と作曲プロセス

2.3.1 節で示した L-System による旋律は、他の多くの例と同様、もともと音符でない対象に音をマッピングする形で旋律を生成したものである。そのため、書き換え規則の音楽的な意義が必ずしも明らかではない。その一端を明らかにするため、作曲家による旋律の作曲プロセスと書き換え規則の関係について述べておく。

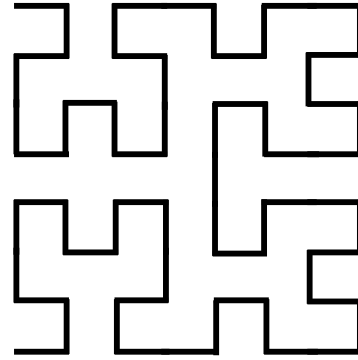
20 世紀の大作作曲家の一人である Olivier Messiaen は、その著作 [1] の中で自らの作曲技法を、自身の作品の例を多数挙げながら体系的に明らかにしている。その中の旋律論に当たる第 8 章は本研究にとって関係が深い。この章で Messiaen は、他の作曲家のイディオムや、民謡、グレゴリオ聖歌等を想起しつつ、これらの中に現れる特徴的な短い音型 (図 2.2) を、自分なりに引用して旋律の構成に用いる方法を解説している。それらの音型は Messiaen 自身の

10 第2章 旋律の形式文法の生成

Initial symbols (axiom): -X
 Rule1: $X \rightarrow -YF+XFX+FY-$
 Rule2: $Y \rightarrow -XF-YFY-FX+$

depth1: -X (axiom)
 depth2: $-YF+XFX+FY-$
 depth3: $-XF-YFY-FX+F+-YF+X$
 $FX+FY-F-YF+XFX+FY-+$
 $F-XF-YFY-FX+-$

A: 文字列の書き換えプロセス



B: グラフィカルな解釈
 (iteration depth 3)



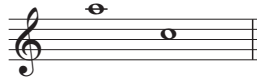
C: 音楽的な解釈 (iteration depth 3)

図 2.1. L-System 文法によるヒルベルト曲線とその音楽的解釈 [4]

作品に例えば図 2.3 のように用いられている。



(a) from harmonic series



(b) from Mozart



(c) from Mussorgsky



(d) from Bartok



(e) from Debussy

図 2.2. Messiaen の引用する短音型の例

これらの譜例では、一つの短い音型が、一つの旋律の中に頻繁に用いられ、その旋律を特徴づけている。さらに、Messiaen はそのような短音型を複数合成してより大きな音型を構成する方法も示しており、実際の作品中での合成音型の使用例を示している (図 2.4)。このような旋律の構成方法は、書き換え規則の用いられ方に通じるところがある。例えば、図 2.4 の合成音型は、図 2.5 (上) に示す三つの書き換え規則 $R_1 \sim R_3$ を用いて図 2.5 (下) のような手順で段階的に生成した結果とみなすことができる。ここで、 R_1 は図 2.2 の (a)、 R_2 は図 2.2

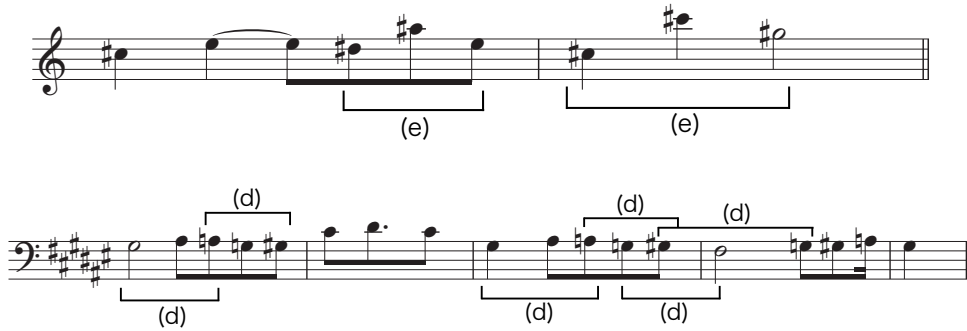


図 2.3. Messiaen の作品中での短音型の使用例. 上: "Arc-en-ciel d'innocence", 下: "Les Mages"

の (d) から作られた規則であり, R_3 は R_1 を反行させて作られた規則である. このように, Messiaen の旋律の構成法は, 書き換え規則の観点から自然に捉え直すことができる. 本研究では, このような見方を全面化し, 旋律法のモデルとする. 少数の短い音型を個々の書き換え規則で表し, それらを次々と適用して旋律を合成していくこのモデルにおいては, 音型の反復を生み出すために特徴を認知しやすい旋律が形成されるとともに, 複数の音型を合成することで機械的な反復を防ぎ, 有機的な旋律の形成が期待できる. そして何よりこのモデルは, 有力な作曲家の実際の作曲プロセスに近いという点からも良いモデルであると考えられる.

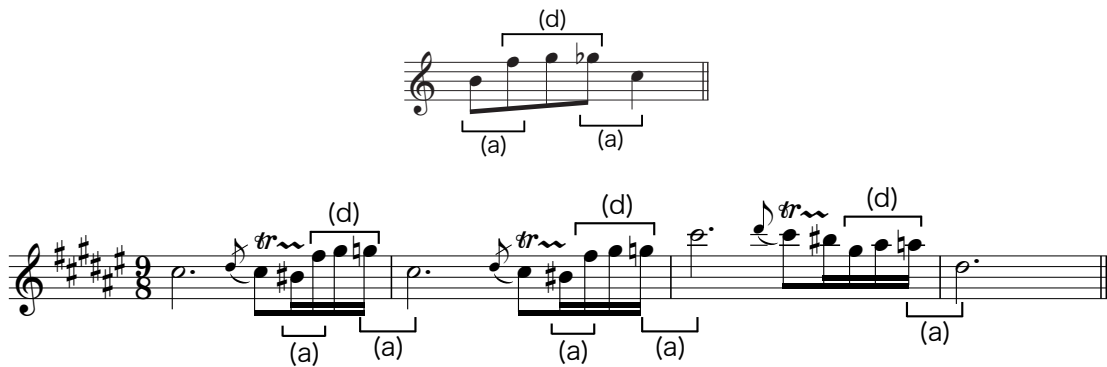
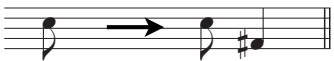



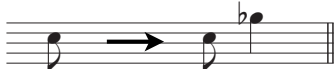
図 2.4. 合成によって形成された音型 (上) と Messiaen の作品 "Chant d'extase dans un paysage triste" におけるその使用 (下)

2.4 多声音楽のための非同期モデル

本節では, 単旋律の生成モデルの多声部への拡張について議論する. 2.4.1 では単旋律の場合の書き換え規則の predecessor と successor の中身をポリフォニックなデータ構造に愚直に

R_1 :  Rule expression of (a)

R_2 :  Rule expression of (d)

R_3 :  The inversion of R_1

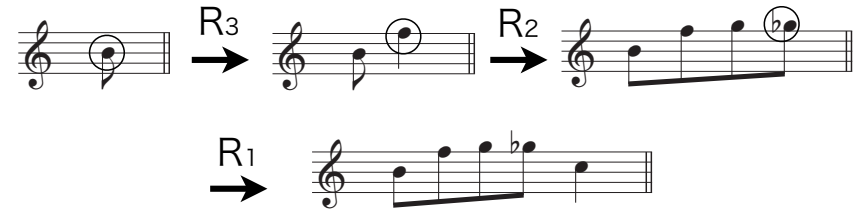


図 2.5. 音型とその合成の書き換え規則による表現

拡張するモデルでは不都合が生じる恐れがあることを示し、2.4.2 では、その不都合を解消するためのモデルを提案する。

2.4.1 単旋律からの拡張の困難

多声部を書き換え規則によって扱う形式は既に [5] 等の研究で触れられている。そこでは次のような表記方法がとられている：

$$(CE)|(GC) \rightarrow D(CE)$$

括弧内の音は同時に演奏される音であり、predecessor 内の “|” は、和音 (CE) の後に和音 (GC) があるというコンテキストを表し、successor 内は単音 D の後に和音 (CE) がくるという意味である。この表記では、括弧内の音を常に同時に鳴らすことになるため、独立したリズムを持った本格的な多声音楽が扱えない。これを愚直に解決するためには、各ノートのピッチの他に長さの情報を導入し、predecessor と successor を次のような類いの形式で表現する必要がある（二声の場合のみ記す）：

$$\left[\begin{array}{l} (p_{11}, d_{11}), (p_{12}, d_{12}), \dots, (p_{1n}, d_{1n}) \\ (p_{21}, d_{21}), (p_{22}, d_{22}), \dots, (p_{2m}, d_{2m}) \end{array} \right]$$

ここで、 p_{ij} は第 i 声部の第 j 音のピッチ、 d_{ij} は第 i 声部の第 j 音の長さである。 p_{ij} は具体的な数値ではなく、休符記号 r または、どのピッチであっても良いという意味のドントケア記号 $\#$ であってもよい。また、両声部の時間の関係として、両方の声部の音価の合計が等しいという条件

$$\sum_{k=1}^n d_{1k} = \sum_{l=1}^m d_{2l}$$

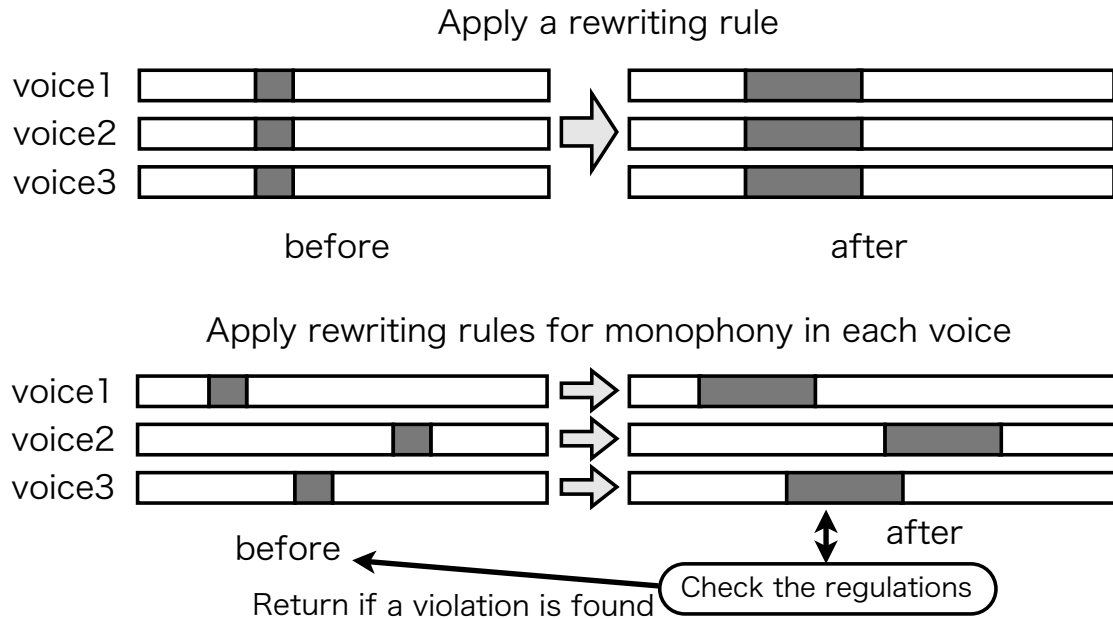


図 2.6. 多声部のための愚直な拡張モデル（上）と提案する非同期モデル（下）

を課すこととする。

ここで問題は、このように愚直に多声部への対応をすると、リズムのバリエーションの増加に伴って書き換えルール中に想定すべき音数が多くなり、必要なルール数が声部数に対して指数関数的に増えることである。もし両声部の音の組み合わせのバリエーションに対応したルール数を確保しなければ、適用できるルールが足りなくなり、それ以上楽曲を成長させることができなくなる恐れがある。それを防ぐため、predecessor 内の音数を制限すると、声部間の組み合わせが同じルールを多用することになり、多声音楽にとって重要な、声部間の独立性が損なわれる恐れがある。また、この表記法では、一つ一つの規則が各声部内の規則どうしの単なる組み合わせの列挙になり、個々のルールの意義と可読性が弱まるという欠点も存在する。

2.4.2 非同期モデルの提案

前節に示した愚直な拡張モデルのような、全声部の情報をまとめて書き換える方法の困難を避けるため、本研究では単旋律の書き換え規則を各声部ごとに別々に非同期に適用する図 2.6（下）のモデルを提案する。また、図 2.6（上）に前節で述べた愚直な拡張モデルを示し、両モデルにおける書き換え規則の適用の違いを示す。ただし、この非同期モデルのように声部ごとに別々に書き換え規則を適用する場合、声部間の調和が取れない場合が出てくる。そこで、それを防ぐため、声部間の関係に対して規制を設け、規制をみたしているかどうかを事後にチェックする重要な手順を導入する。ここでいう規制とは、多声音楽としての声部間の良い関

14 第2章 旋律の形式文法の生成

係を保つための条件をルールとして事前に手動で与えるものとする（自動生成の対象はあくまで個々の旋律のスタイルである）。ただし本研究では、旋律の形を第一義に考える焦点化のため、和声に関する規制は設けないこととし、それについては今後の課題とする。具体的に本研究では、次の4つの規制を設定する：

- 声部の上下関係を明確にするため交叉を禁止する
- 全ての声部に共通した音域の上限、下限を設ける
- 拍節構造を保つため、小節の冒頭では過半数の声部で発音しなければならない
- 声部間の独立性を保つため、同時に同じリズムを用いることを禁止する

この提案モデルのもとでは、初期値 (Axiom) および書き換え規則の集合を与えた上で、次の3ステップを、所望の長さになるまで反復 (T 回とする) することで楽曲が生成される：

T times iteration {

step 1: 各声部に対して、ひとつずつ、書き換え規則を適用する対象をランダムに選ぶ。

step 2: 各声部に1回ずつ、1で選択した対象に書き換え規則を適用し、各声部を同じ長さに成長させる。

step 3: 延長した旋律に対し、旋律間の関係性の規則をチェックし、違反がなければ終了。もし違反があれば、step1からやり直し。

}

ここで、全ての声部の旋律の長さをそろえるためすべての書き換え規則において、successorの長さは predecessorの長さよりも L だけ長いものとして単純化して考えることとする。またここで、後のため他の記号の定義も与えておく。 V を声部数とし、各声部の書き換え規則の数は R 個で共通とする。また、 C_i を第 i 声部に対する単旋律の書き換え規則の集合 ($1 \leq i \leq V$)、 $c_{ij} (\in C_i)$ を第 i 声部の第 j ルールとする ($1 \leq j \leq R$)。 C_i は声部間で共通にすることもでき、別々にして「複数スタイル」の楽曲とすることもできる。上記の反復プロセスが終了すると、(最初の長さ + $L \times T$) の長さの楽曲が得られる。このモデルは、声部間の独立性を保ち、声部間の組み合わせの可能性を制限しすぎないため、多声音楽にとってふさわしいと考えられる。愚直な拡張モデルにおいては、書き換えルールに事前に声部間の規制が反映されることとなるが、提案する手法では、書き換え規則の適用後にチェックすることになる。書き換え規則に声部間の規制を織り込むことは、規則の適用が容易である反面、規制できる範囲が predecessorの内部に限定されるという欠点が存在する。適用後のチェックは楽曲全体を範囲として行うこ

とが可能であり、ローカルな範囲しか規制できない書き換えルールそのものの弱点を補うことができる。また声部数に対してルールの数は $V \times R$ と線型になり、ルール数の肥大化をおさえられる。

なお、複数の声部を並列的に生成するのではなく、一声ずつ順番に生成する方法も考えられるが、その場合、後に作曲する旋律ほど規制を強く受けるという声部間の不平等を生むため、ここでは同時に成長させるモデルを採用した。

2.4.3 より発展的な対位法的形式の扱い

多声音楽は、逆行、反行、拡大、縮小などの音型操作や、カノン、フーガ等の摸倣による形式との関係が強い。それらの実現は次節以降で本研究が扱う範囲を超えるが、提案モデルの中でそれらを扱える可能性があるかどうかは重要な問題であるため、ここではその見通しについて触れておく。

まず、逆行および反行については successor の音型を逆行、反行させた書き換えルールによって扱うことが考えられる。拡大、縮小については、書き換えルールの1回の適用で L だけ旋律を延ばすという前述の制限の中では長さが合わず、通常のやり方では扱えない。ただし、例えば2倍の拡大については、他の声部で2回ルールを適用するときに successor を2倍に拡大したルールを1回適用することで声部間の拍数を合わせる、といった形で対応することができる。2分の1の縮小については、他の声部にルールを1回適用する際に縮小のルールを2回適用することで声部間の長さを合わせるができる。他の倍率の拡大、縮小についても同様である。また、逆行、反行、拡大、縮小のルールを適用した後、その形を保つためには、後からその部分に対してさらに他のルールを適用して音型を壊してしまうことを禁じる措置が必要となろう。

カノンについては、声部ごとのルールの共有に加えて、複数の声部においてルールの適用を対応する位置に同時に指定することで実現できる可能性がある。ただし、各声部でのルールの適用位置が対応する箇所でなければならないという制限が加わるため、声部間の規制を満たすのが困難となる恐れもある。フーガについては、ルールの合成によって一声の長い主題を始めておき、次にそれに合う対主題を、転回可能という条件や、主題と対主題の対比に関する条件を声部間の規制に追加して作るというやり方が考えられる。ここで、主題、対主題はルールの合成によって作られた、拡大と同様の新たな長いルールとして扱うことができる。ただし、この主題と対主題の形成ルールは同期的に適用しなければ共存することが難しいと考えられる。断片化された主題が出現するストレッチについては、主題形成ルールの合成を途中でやめたルールによって表現でき、曲の後半でのみ使用するなどの制限をつけて用いることができる。

以上のように、提案モデルはこれらの形式を統合できる可能性を持っている。ただし、これらを扱いながら実際に曲を生成できるルール集合を獲得するには、次節で述べる、扱わない場合のアルゴリズムよりも制約が厳しく、より解を見いだすのが困難となる恐れがある。本研究ではその問題については扱わず、実現の可能性と困難の指摘にとどめる。

2.5 旋律文法の生成手法

前節では、多声音楽を扱うための一般的な文法モデルを提示した。次の問題は、具体的な文法規則を自動生成することである。本節では、ルール集合を獲得するための遺伝的な機械学習方法として知られるクラシファイアシステム (CS[17, 18]) を応用した旋律文法の生成手法を提案する。

2.5.1 クラシファイアシステム

クラシファイアシステムとは、遺伝アルゴリズム (GA) の研究で有名な John Holland によって提唱された適応的システムである。システムは、クラシファイアと呼ばれる手持ちのルールの集合に基づいて行動を起こし、行動が環境に対して有効な働きかけをした場合に環境から報酬を受け取ることで、自らに有利な規則を学習、獲得し、環境に適応していく。さらに、信頼度の低いクラシファイアは GA によって生成される新しいクラシファイアと入れ替えられ、淘汰される。これにより、価値の高いルール集合が創発されていくことが期待できる。図 2.7 にその概念図を示す。

クラシファイアは、if 〈 condition 〉 then 〈 action 〉 という形式のルールである。環境からの情報が condition 部分を満たしている時に、システムは対応する action を環境に送り出す。各クラシファイアには信頼度 (credit) と呼ばれる数値が割り当てられており、報酬をもたらしたクラシファイアの信頼度を相対的に増加させるアルゴリズムによって随時更新される。条件部に当てはまるクラシファイアが複数ある時には、信頼度の高さに応じて確率的に選択される。本研究では、クラシファイアの condition と action に書き換え規則の predecessor と successor を対応させ、環境を生成中の楽曲に対応させることができるため、CS は目的にかなった手法だと考えられる。

2.5.2 文法規則の信頼度の評価方法

本研究において CS を応用するには、旋律法としての望ましさを数値的に評価し、その値を信頼度に反映する必要がある。評価方法を設定する上では次のことを考慮する必要がある：

- (1) 少数の規則を多用することで旋律の特徴が際立つと考えられる。
- (2) しかし、規則が少ないと適用できる規則が無くなりやすく、曲を長く成長させるのが困難になる。そのため適用可能なルールや他のルールの適用を可能にするのに役立つルールは高く評価すべきであろう。そのようなルールは、曲の成長への貢献度 (使用された頻度) を基準とすることで評価できると考えられる。
- (3) 規則の形に関する人為的な方向付けをしすぎると、旋律スタイルの自動生成の目的に反する恐れがある。

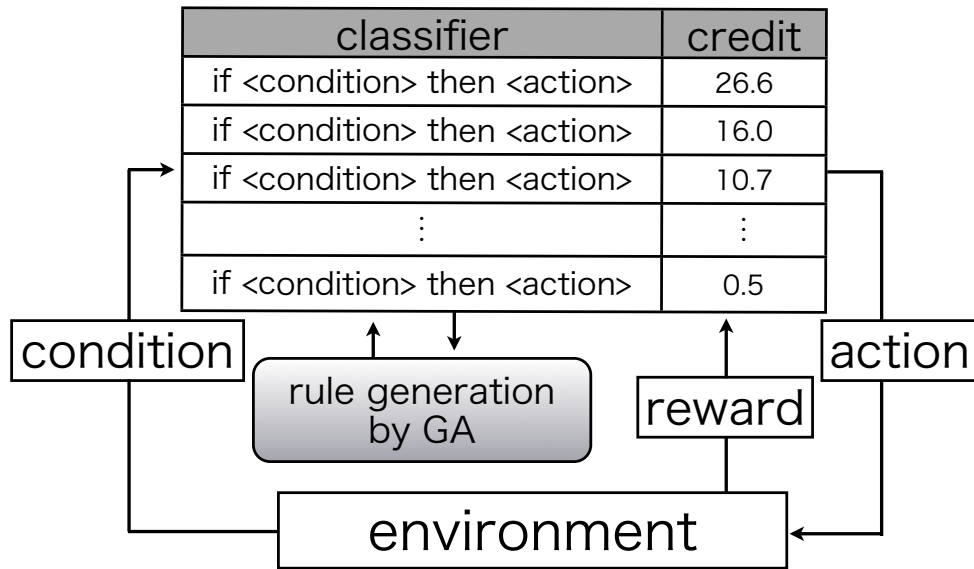


図 2.7. クラシファイアシステムの概念図

以上の点をふまえ、次のような観点からルールを獲得に臨み、ルールの信頼度へのフィードバックを設定する。

- 各声部のクラシファイアの数 R をできるだけ小さくおさえる（アルゴリズムを遂行する前に固定する）。
- 個々のルールが曲の成長に貢献した度合い（そのルールが用いられた頻度）を報酬として、そのルールの信頼度に割り当てる。
- 規則の形を方向付ける人為的な評価基準はあえて導入しない。

2.5.3 非同期モデルへの適用

本節では、2.4.2 節で提示した多声音楽の非同期モデルに対して CS を適用し、書き換え規則の集合を生成する具体的手順を述べる。この手順は次の 5 ステップからなる：

- step 1: 初期値を設定.
- step 2: 書き換えルールの反復適用による作曲 (規定の反復回数内に所望の長さまで曲を延長できた場合, プロセスを終了).
- step 3: 信頼度の更新.
- step 4: 成長した長さが前回の最大値以上だった場合, そのルール集合および信頼度を保存しておく. そうでなければ前回まで保存されていたものに戻す.
- step 5: 遺伝アルゴリズムでルール集合を更新し, step2 に戻る.

各ステップの詳細は以下の通りである. step1 ではシードとなる楽譜の初期状態を与え, 書き換えルール集合 C_i の初期値をランダムに与え, 各書き換えルールの信頼度を初期化する. step2 では, 前節のステップ1から3のステップによって曲を一段階ずつ成長させる. 各成長段階で規則の適用のトライアル回数の上限を設け, 上限を超えたら手持ちのルール集合では曲の成長がこれ以上不可能であると判断し, step3 に移る. 各成長段階におけるルールの選択は, 各声部ごとに一回ずつ, 信頼度によるランキング選択を行う. step2 においてルールの選択, 適用を繰り返した結果, 曲の長さが所望の長さに達した場合, 全てのプロセスが終了となる. この時, 求める旋律文法とそれに基づく楽曲が同時に生成されたことになる. step3 では, 各ルールの善し悪しを評価する. 具体的には, step2 での曲の成長に寄与した規則を, 規則の使用頻度に応じて評価する. これによって, 曲の成長を持続可能なものにし, 目標となる楽曲の長さに近づけて行く. 各規則の信頼度は, (その規則が今回使用された回数+前回まで保存されていた信頼度 $\times \gamma$) で更新する. ここで γ は0と1の値をとる実数値であり, 前回までの信頼度のある程度割り引いて保存しつつ信頼度の発散を防ぐためのものである. step4 では, 過去に最も良い結果を残したルールセットを保存しておく. これは, GA でのルールの入れ替えによって, 結果が悪くなることを避けるための処置である. step5 では, step3 での信頼度評価に基づき, 信頼度のランキングの低い規則を, 遺伝的演算によって生成された新しいルールと置換し, 規則の集合をアップデートする.

遺伝的演算には表 2.1 のような交叉と突然変異を用いる. ただし単純化のため predecessor の音数は一音とする. 図 2.8 に一点交叉の手順を例示する. step1 で定める初期値については, 様々な設定法が考えられるが, 本研究では次のように設定する. シードの楽譜の初期状態は, 各声部ごとに長さ L の 1 音とし, 各声部の初期値固有に設定した音域内からランダムな音高を選択する. 声部間の上下関係を守るため音域は重複させない. C_i の初期値は, predecessor の長さ, successor 内のオンセット数, オンセット位置, successor 内の各 2 音間の音程の順に可能な範囲内でランダムに選択する. ただし音程は, 上下ともに 1 オクターブを限度とし, 順次進行の確率のみ他の音程よりも高く設定する. 各書き換えルールの信頼度は

交叉	
successor を一点交叉させ, successor の音価の変化に応じて predecessor の音価を伸縮調節する	
突然変異	
1	successor の内部で音高を一音変化させる
2	successor の内部で二音を入れ替える
3	successor の内部で二音の音高を入れ替える
4	successor の内部で二音の音価を入れ替える
5	successor の内部で一音削除して削除した音の分の音価をいずれかの音に分配する
6	successor の内部で音数を一音の音価を二つに分割して新しい一音を挿入する
7	predecessor の音価を変化させ, successor のいずれかの音に変化した分の音価を分配する
8	successor のいずれかの音を削除し, predecessor の音価をそのぶん減らす
9	successor に音を挿入し, predecessor の音価をそのぶん増加させる
10	ランダムにルールを生成する

表 2.1. GA に用いる遺伝的演算のリスト

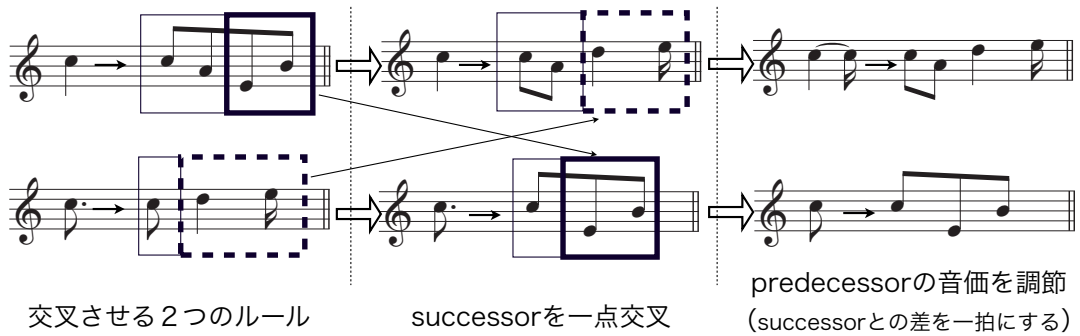


図 2.8. 一点交叉の例

1.0 として初期化する.

2.5.4 楽曲の生成と観察

本節では、提案手法による旋律文法と楽曲の生成実験について述べ、生成楽曲サンプルを紹介する。用いたコンピュータ環境は Mac OS 10.6.8, 2.4GHz Intel Core 2 Duo であり、プログラミング言語には Ruby を用いた。

生成の条件は次のように設定した。声部数は3声とし ($V = 3$)、各声部の規則の数は7とした ($R = 7$)。旋法は、Messiaen が理論化し、自作に用いたことで有名な、移調の限られた旋法の第2番 (MLT2 と略記する) を用いた。声部数は3声 ($V = 3$)、拍子は4分の2拍子、曲の長さは50拍、音価の最小単位は16分音符、音域の上限および下限はMIDI番号で96および36として統一した。書き換え規則の形式は次のように限定した。successor のトータルの長さは predecessor に対して1拍分長くなるものとし、successor の開始音の音高は predecessor の音と同じとする。休符は用いないものとし、各声部の音符の長さは、声部ごとに別々に設定した最小単位の倍数をとるものとする。これによってポリリズムが可能である。ピッチや書き換えルールは音階に基づくものとする。書き換えルールとしては、自動生成する通常のもの他に、終止形を形成するルールを手動で設定した。このルールは、任意の音階音から上または下に順次進行し、二分音符の終止音を追加するものであり、任意の音階音が終止音となりうる。終止形ルールは一拍ごとの成長段階において10回に1回適用し、旋律に区切りを与えるために用いた。終止型ルールの適用方法は基本的に通常の書き換え規則と同様である。ただし、最終小節のみ全声部同期した終止型を付け加えた。ルール更新におけるGAの世代間で、保存するルールの割合は5/7とし、それ以外の2/7をGAの交叉または突然変異によって置換した。交叉と突然変異の比率は7:3とした。

図2.9および図2.12に生成された楽曲の二つのタイプの例を示す。図2.9の曲は声部間でルールを共有した場合（それに伴い音価の最小単位も共有）であり、図2.12の曲は声部ごとに異なるルールをもつ場合である。後者では、音価の最小単位も声部ごとに異なるように設定した。両例とも生成には数分を要した。まず、両例を観察すると、提案モデルにおける声部間の制約の効果を明確に確認することができる。それぞれの声部は音域の衝突を回避し、リズム的な独立や拍節構造が確保されていることがわかる。

次に、両例を個別に観察する。図2.10はこの楽曲を生成した文法である。この曲で特に目立つのは R_3 から R_6 にみられるような順次進行による16分音符の高速の下降や、 R_4 から R_7 にみられるような上方への6度の跳躍 (MLT2内では5度ではなく6度)、 R_1, R_2 にみられるような3度の下降等を含む音型であり、声部間でのルールの共有により各声部で反復され、対位法音楽における摸倣のような声部間の呼応関係を生じさせていることが確認できる。他方、図2.12の曲は、各声部でルールを別々に設定した「複数スタイル」の楽曲の生成例である。上声は非常に細かい音価で激しく動いているが、スイングのような付点のリズムが特徴的な旋律スタイルとなっている。中声は、最小音符は一拍の6分の1と細かく設定されていたが、細かい音価は出現せず、一拍を2分割する音価による、緩やかなリズムをとっている。下声部は一拍を三分割するリズムに基づいており、中声部とは2:3のポリリズムの関係になっ

The image displays a musical score for Example 1, consisting of five systems of three staves each. The music is written in 2/4 time. The first system (measures 1-5) shows a treble clef staff with a melodic line and two bass clef staves with accompaniment. The second system (measures 6-10) continues the melodic development. The third system (measures 11-15) features more complex rhythmic patterns. The fourth system (measures 16-20) shows further melodic and harmonic evolution. The fifth system (measures 21-25) concludes the piece with a final cadence. The notation includes various accidentals (sharps, flats, naturals) and articulation marks (accents, slurs).

図 2.9. 生成楽曲例 1. 書き換え規則および音価の最小単位は声部間で共通.

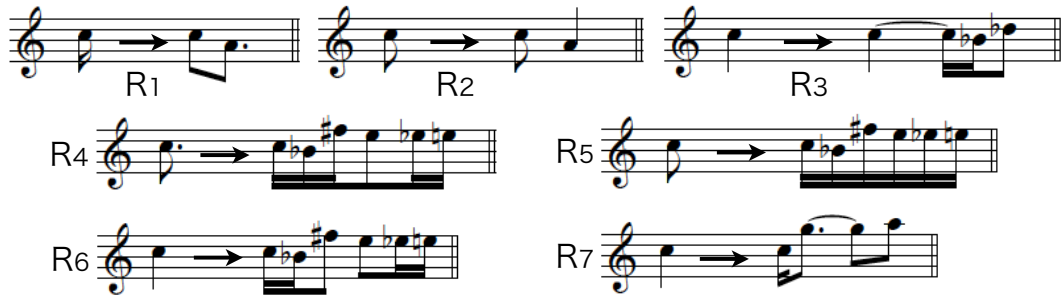


図 2.10. 図 2.9 の楽曲の生成ルール

ている。それに加えて、上声部の細かい動きとの対比も感じさせ、全体として三つの旋律スタイルが共存している。この曲の上声部で扱われているような非常に複雑なリズムの領域は、古典的なポリフォニー音楽で扱われないため範例が少なく、コンピュータによる統計的な学習を方法論として自動作曲を行うことは困難である。このような領域に踏み込める点にも、本手法には意義があると考えられる。

図 2.11 に図 2.12 の楽曲の生成過程での、世代ごとの成長した曲の長さ (拍数) を図示する。遺伝アルゴリズムの初期世代においては、曲をほとんど成長させることができていないが、世代を経るごとに有効なルールを獲得し、成長の可能性が増加している。また、参考までに声部ごとに異なる旋法を用いた場合の楽曲生成例を図 2.13、図 2.14 に示す。

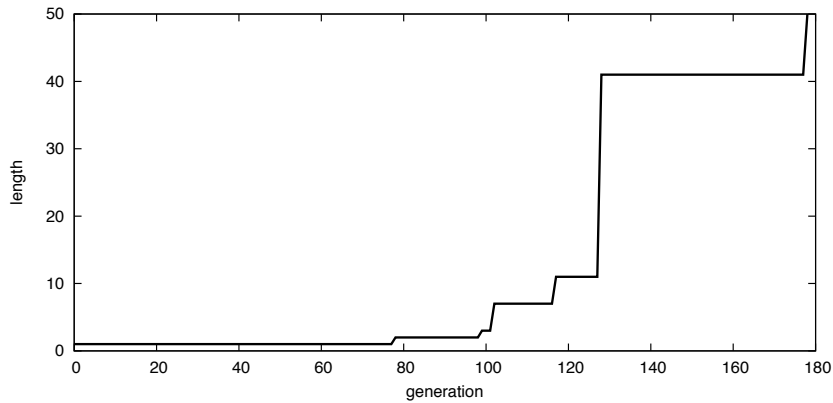


図 2.11. 図 2.12 の楽曲生成における成長曲線

The musical score consists of four systems, each with three staves (treble, middle, and bass clefs). The time signature is 2/4. The score is characterized by complex rhythmic patterns, primarily using 9-measure phrases and 3-measure phrases. The first system (measures 1-6) shows a dense texture with many 9-measure phrases in the upper staves and 3-measure phrases in the lower staves. The second system (measures 7-12) continues this pattern, with the upper staves featuring more complex rhythmic structures. The third system (measures 13-19) shows a similar pattern, with the lower staves featuring more complex rhythmic structures. The fourth system (measures 20-26) concludes the piece with a final 9-measure phrase in the upper staves and 3-measure phrases in the lower staves.

図 2.12. 生成楽曲例 2. 書き換え規則および音価の最小単位は声部ごとに異なる。

The image shows a musical score for Example 2.13, a three-part setting in 2/4 time. The score is divided into four systems, each with three staves (treble, alto, and bass clefs). The music features complex rhythmic patterns and various accidentals (flats, sharps, naturals) across the different parts. Measure numbers 8, 14, and 20 are indicated at the start of their respective systems.

図 2.13. 生成楽曲例 3. 各声部ごとに異なる旋法を使用. 音価の最小単位は各声部に共通.

図 2.14. 生成楽曲例 4. 各声部ごとに異なる旋法を使用. 書き換え規則および音価の最小単位も声部ごとに異なる.

2.6 生成曲の評価実験

提案手法が旋律のスタイルと多声音楽の生成にどのような有効性をもつのかを調べるため、多数の楽曲を生成し、作曲家（作曲を専攻する学生も含む）による評価実験を行った。作曲家を被験者とするのは、新しいスタイルを評価するには、もし専門でない人が評価した場合は単になじみ深いかどうかによって判断がなされ、新しい音楽が理解されないなどといったことが起こる恐れがあるが、作曲の専門家であれば比較的正しい判断ができると考えたためである。

2.6.1 評価基準

評価基準の設定は、旋律スタイルの評価のため「旋律スタイルの一貫性」、「旋律スタイルの独創性」、「旋律の動きの豊かさ」の3つの基準、また多声音楽としての声部間の関係の良さの評価のため「声部間の調和」、「声部間の独立性」の2つの基準、それに加えて「楽曲の総合的な良さ」を設定した。これら6項目を7段階で評価する。7段階の意味は、7：極めて高い、6：高い、5：やや高い、4：どちらでもない、3：やや低い、2：低い、1：極めて低い、とする。

評価楽曲の種類の設定においては、文法ルール数 R および、各声部でルールを共有するか否か（このパラメータを S とし、 $S = 1$ を共有、 $S = 0$ を非共有とする。）の2つ要因を設定し、パラメータによる評価の変化を調べる。 R についてはモデルを立てる時点で、 R が小さいことが旋律スタイルにとって重要との予想をしていた。それを実証するために R による評価の変化を調べる必要がある。また、声部ごとに異なる旋律スタイルをとる場合と、全ての声部が同じスタイルの場合の違いを調べることは、多声音楽と旋律スタイルの関連性をみる上で最も大きな差異だと考えられるため、 S の違いを評価対象に含める。

2.6.2 実験条件

実験参加者は、作曲家あるいは作曲を専攻する学生26名（男性15名、女性11名、年齢：20～41）であった。実験に用いた楽曲は、 R について小中大の3水準（7,14,21）、それぞれの R について $S = 1$ および $S = 0$ の2水準、合計 $3 \times 2 = 6$ 通りの条件設定で、それぞれ5曲ずつ、計30曲を生成したものを用いた。すべての曲は約40秒のMIDIファイルであり、聴取順序の影響を相殺するために実験参加者ごとの聴取順序はランダム化した。楽曲の生成の条件設定については、 $S = 0$ のときを含め、音価の最小単位を16分音符に固定した。このことと R を可変にしたことと以外は、節に記述した条件と同じである。声部数や旋法の違い、声部ごとに異なる旋法を持つポリモーダルな楽曲、声部ごとに音価の最小音符が異なるポリリズム的な楽曲等を考えると、組み合わせの数が膨大になる。評価者が評価できる分量の限界を考慮してこれらは対象外とした。

2.6.3 実験結果

得られた6つの評価値の平均値を折れ線グラフにプロットし、図 2.15 に示す。まずこの表から、 $R = 7$ のときは全ての評価項目においてニュートラルな 4 の値を上回っており、 R が小であることが提案モデルにとって重要であることが把握できる。ただし、すべての項目において平均値は 4 から 5 付近に位置しており、その効果は非常に顕著というわけではないことがわかる。次に、効果を詳細に分析するため、各評価値に対して R と S を要因とした 2 元配置分散分析を行った。その結果、全ての項目において交互作用は現れなかった。 S に関する主効果は「旋律の動きの豊かさ」において現れ、 R に関する主効果は「旋律スタイルの一貫性」、「旋律スタイルの独創性」、「声部間の調和」、「楽曲の総合的な良さ」において見いだされた。「声部間の独立性」においてのみ、いずれの主効果もみられなかった。

主効果のみられた項目について多重比較 (S に関しては t 検定) を行ったところ、表 2.2 のような結果となった。ここで $R = 7 > R = 14$ という表記は $R = 7$ のときの方が $R = 14$

評価項目	結果	検定方法	p 値
旋律スタイルの一貫性	$R = 7 > R = 14$	TukeyHSD	0.031
	$R = 7 > R = 21$		< 0.001
旋律スタイルの独創性	$R = 7 > R = 14$	TukeyHSD	0.030
	$R = 7 > R = 21$		0.015
旋律の動きの豊かさ	$S = 0 > S = 1$	t 検定	< 0.001
声部間の調和	$R = 7 > R = 14$	Dunnnett	0.045
	$R = 7 > R = 21$		0.001
楽曲の総合的な良さ	$R = 7 > R = 14$	TukeyHSD	< 0.001
	$R = 7 > R = 21$		< 0.001

表 2.2. 多重比較の結果

のときよりも高評価であることを表すものとする。多重比較の検定方法は基本的に Tukey の HSD 法を用いたが、有意差がみられなかった「声部間の調和」においては、より焦点を絞った多重比較法である Dunnnett 法を試みた。この結果から「旋律の動きの豊かさ」に関しては、 $S = 0$ のときの方が $S = 1$ のときよりも高評価だったことがわかる。このことは、複数の旋律スタイルの共存が、一定のポジティブな効果をもつこと示唆する。また、 R に関する分析から「旋律スタイルの一貫性」、「旋律スタイルの独創性」、「声部間の調和」、「楽曲の総合的な良さ」についてはいずれも $R = 7$ (小) のときに最大となることがわかった。このことは、 R が

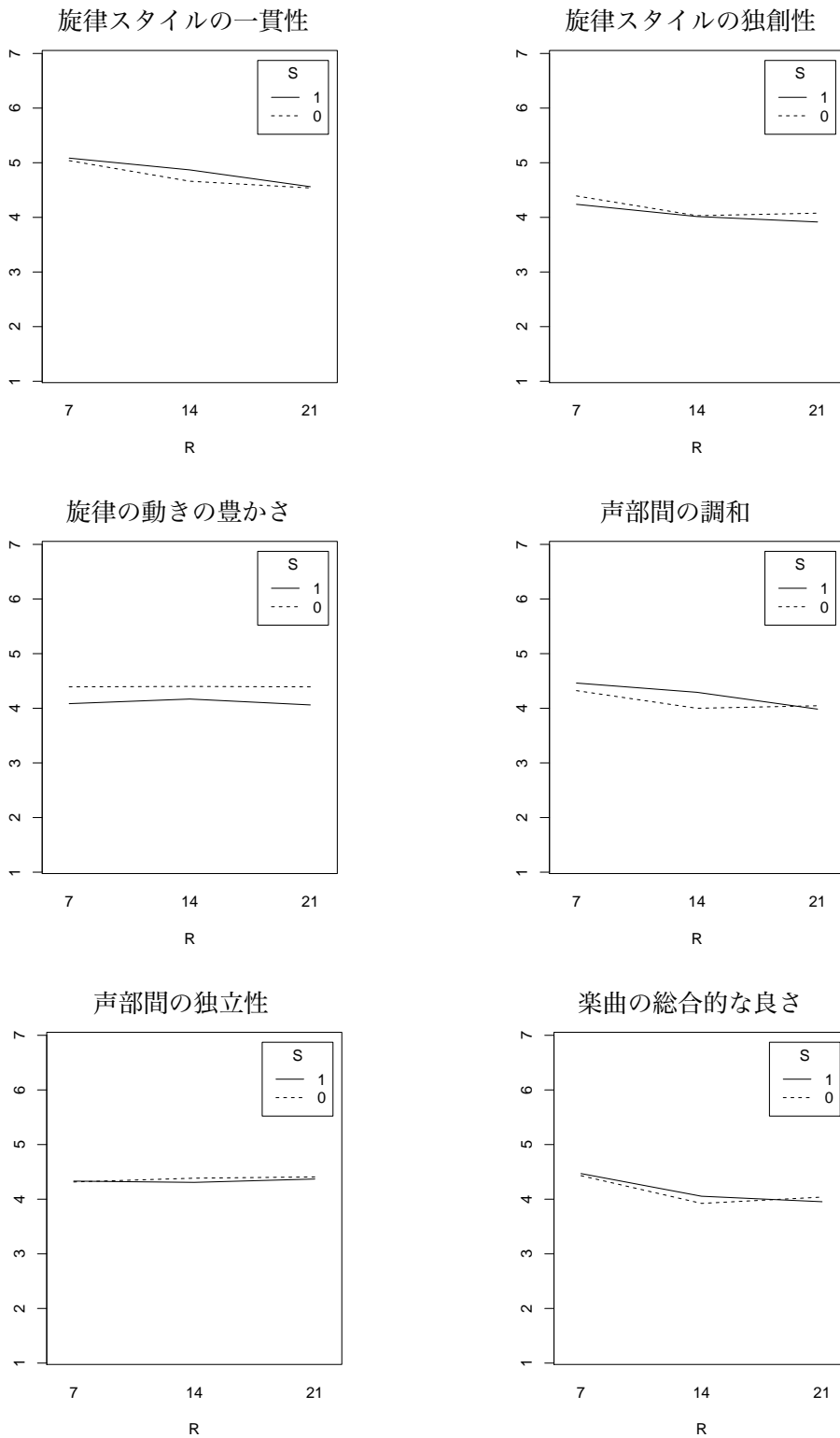


図 2.15. 各評価項目の平均値

小さい方が旋律スタイルが向上するという当初の予想を裏づける結果である。

また、本実験では S について主効果がみいだせなかった項目が多いが、今後サンプル数を増やす等、より正確な検証の価値がありそうな仮説として以下のようなものがある：

仮説 1: 「旋律スタイルの一貫性」は共有のときの方が評価が高い。

仮説 2: 「旋律スタイルの独創性」は非共有のときの方が評価が高い。

仮説 3: 「声部間の調和」は共有のときの方が評価が高い。

仮説 4: 「声部間の独立性」は非共有のときの方が評価が高い。

仮説 1 はグラフ上では常に成り立っており、ルールの声部間での統一がスタイルの認知に繋がりを、一貫性を高めると考えれば自然だと考えられる。仮説 2 もグラフ上では常に成り立っている。仮説 3 は、声部間でのルールの統一は声部同士の関係を強め、調和を高めると考えれば意味として自然であり、グラフ上では $R = 7, 14$ のとき成り立っている、さらに $R = 14$ のときには t 検定で有意差が確認できている ($p=0.034$)。仮説 4 に関して、声部ごとに異なるルールをもつと声部の独立性が高まることは意味として自然であり、グラフ上では $R = 14, 21$ のとき成り立っている。

2.7 章のまとめ

本研究では、多声音楽のための旋律法を自動生成する手法を提案した。そこではまず、複数の声部の旋律を非同期的に書き換え、成長させるような一般的な文法モデルを構築し、次に具体的な旋律文法規則の集合をクラシファイアシステムによって獲得し、旋律スタイルを生成する手法を提案した。また生成実験により、実際に文法規則と楽曲を生成が可能であることを確認したのち、多数の生成曲を作曲の専門家によって評価する実験を行った。その結果、非常に顕著な効果というわけではないものの、文法ルールの数を少なく設定することが多くの評価項目を高めることが確認された。それらは、旋律スタイルの一貫性と独創性、声部間の調和、曲の総合評価の 4 項目である。このことは、少数のルールから旋律を生成するという、モデルを考案した当初の意図が支持されたことをも意味する。また本手法は、異なる旋律スタイルが共存する「複数スタイル」の楽曲を生成できるが、そのような生成曲は、声部間で同じルールを共有する場合よりも「旋律の動きの豊さ」を若干増す効果があることが確認できた。この評価実験の結果は、生成される作品のスタイルをコントロールするのに用いることができる。

今後のさらなる発展的な課題としては、2.4 節で述べたモチーフ操作やカノン、フーガ等の形式を扱えるように本手法を拡張することが挙げられる。また本研究は、楽曲のメタレベルのスタイルをも生成するというより深いレベルから楽曲を生成する研究の一環である。したがって、旋律スタイルのみならず、楽曲のマクロな形式や、旋律間の関係性を規制する声部間的な規則（対位法や和声等）、さらに旋律法の基礎となる旋法（本研究では旋法は予め与えられていた）をも自動生成の対象に含め、楽曲スタイルをより包括的に自動生成できるようにしていくことが重要と考えられる。また、旋律文法ルールの可読性を生かして人間のユーザが文法を部分的に指定し、それと整合する他の規則の集合を生成したり、提示された文法をユーザが好

30 第2章 旋律の形式文法の生成

きなように修正した後，楽曲を再生成するようなインタラクティブな作曲支援システムに本手法を応用することも考えられる。

第3章

感情を表現する旋法の生成

3.1 旋法と感情

本章では、音楽理論の最も基礎的な部分の1つである旋法を生成対象として扱う。旋法には多くの種類が存在し、どれを用いるかによって音楽の表現内容が大きく変化する。そして、新たな表現力をもつ旋法を生成することは新たな音楽スタイルを生み出すことにつながる。また前章で扱った旋律法もその基礎に旋法があり、旋法は旋律法の表現力にも大きな影響を与える。

では、どのような性質を持つ旋法を生成すべきだろうか。それを考える上で重要になるのは旋法と感情との結びつきである。音楽は感情の表現と関係が深い。長旋法の明るさと短旋法の暗さからもわかるように、とりわけ旋法は感情との結びつきが強い。その関係性については古代ギリシャの時代から指摘されてきた。例えばプラトンの「国家」[20]においてソクラテスは、国家の教育的な観点から、悲しみを帯びた混合リディア旋法や高音リディア旋法、柔弱なイオニア旋法等は排除すべきであり、勇気と節度を表すドリア旋法とフリギア旋法を用いるべきだと論じている¹。また、アリストテレスは「政治学」[21]において、教育目的の倫理的な旋法としては、落ち着きと男性性を最も良く表しているドリア旋法がふさわしいが、場合によっては行動的な旋法や熱狂的な旋法にも使い道があると論じている。このように、特定の感情を表現することが旋法にとっての重要な役割だと考えられる。

ただし、ある感情を最も良く表現する旋法は何かということを実験によって明らかにすることは難しい。なぜなら、存在する旋法の種類の膨大さゆえ、すべての旋法の可能性に対してひとつひとつ聴取実験を行うことは現実的でないからである。

そこで本研究では発想を転換し、予め用意された旋法を評価する心理学的なアプローチではなく、強化学習を用いた機械学習的なアプローチを用い、与えられた感情に対する表現力の高い旋法を獲得する手法を提案する。この手法は、人間の評価者の感情に旋法の方を適応変化させていくため、あらゆる可能な旋法について調べることなく目的の旋法に到達できる可能性がある。また、長短の旋法や教会旋法等、現在までに使用されてきた旋法が使われるようになって

¹ ここでの旋法名は教会旋法と同一名称の異なる旋法である。

た経緯は、人間の試行錯誤による動的、歴史的な過程であると考えられ、成功や失敗の経験による学習を模擬した強化学習の手法を用いることは自然であろう。

ただし、この方法においては人間の評価者にはそれなりの負担がかかる恐れがある。それを軽減するための工夫として、パラメータ数を減らして効率的に学習できるようにするための、価値関数の関数近似モデルを構築する。

もし旋法と感情の対応づけができるようになれば、ドラマや映画や劇などのシーンに適した感情をもつ楽曲の生成や、パブリックスペースや自宅などでの様々な場面に合った雰囲気をもつBGMの生成など、楽曲の提供のための重要な要素技術となるだろう。また、音楽療法において、患者の感情に働きかけることで、望ましい心理状態に患者を誘導していくような応用の可能性も考えられる。さらに、12音以外の微分音による調律システムなどのマイナーな調律システムにおいてどのような旋法を使えば良いかという問題に解決を与えることもできるかもしれない。

3.2 関連研究

旋法と感情の対応についての実証的な研究は、ここ数十年さかんに行われており、とりわけ長短の旋法に関する研究結果は多数存在する。Kastnerらは、幼い児童が旋法による感情の違いを判別できるかどうかを調べるため、感情を表す顔のイラストをラベルに用いた実験 [22] を行い、3から12歳までの38名の被験者の児童全員が、長旋法の旋律と肯定的な感情および、短旋法の旋律と否定的な感情の関係を感じとれたと報告している。長短の旋法のみならず、それらを含む7種類の教会旋法を用いれば、より多様な感情を表現することができる。Hillらは、バッハによるイオニア旋法とフリギア旋法の曲を被験者に聴かせ、救済 (salvation) と断罪 (condemnation) を表すのはどちらか判定させるという実験を行い、イオニア旋法が救済、フリギア旋法が断罪に対応するという結果を得ている [23]。また、Ramosらは7種類の教会旋法を被験者実験から、快不快 (valence) と覚醒度 (arousal) による感情の2次元モデルにマッピングし、それぞれの旋法の表す感情の違いを示した [24]。

西洋以外でも、インドの旋法であるラーガ (サンスクリット語で心や感情を彩ることを意味する) においては、多数の旋法を使い分けることで、様々な気分を表現できるといわれている [25]。ラーガは単なるピッチクラスの集合としての音階ではなく、その中での旋律的なふるまいをも含んだ概念であり、音楽辞典 [26] においては、一種の「旋法 (mode)」として扱われている。本研究では、旋律的なふるまいをも研究対象に含め、「音階」ではなく「旋法」という用語を用いる。

旋法と感情の関係を研究する上で大きな障害となる問題は、旋法の種類が数学的に膨大にあるにもかかわらず、現実的には限られた数の旋法についてしか調べることができないということである。単純計算すれば、1オクターブ n 音の音律において、旋法の種類はピッチクラスの違いのみを考慮しても 2^n (12音では4096) 個存在する。実際、多くの研究は長短の2つの旋法しか扱っておらず、[24]のように7種類の教会旋法を調べるだけでも意欲的であろう。また、ピッチクラスの集合としての音階だけではなく、どの音からどの音へ遷移するかという情

報を含めた旋法を対象とすれば、旋法のバリエーションはほとんど無限である。

そこで本研究では、心理学的な実験ではなく、強化学習 [27] を用いた工学的アプローチでこの困難を回避する手法を提案する。それは、目的となる特定の感情に対して人間からの感情の評価のフィードバックを用いて、旋法の方を感情に適合するように適応変化させるというものである。これにより、固定した旋法を無数に用意してそれぞれを評価するという作業を行わずに、目的の旋法を獲得できる可能性がある。また、この方法は、個人差や文化の違いの問題を回避できるという利点がある。

強化学習を音楽へ応用する既存研究としては、人間の望む音楽的緊張感の高低に対して、音楽エージェントのパラメータを自律的に適応変化させる研究 [28] がある。エージェントがコントロールできるパラメータはテンポ、音量、アーティキュレーションの三つであり、エージェントが実際に適応変化しうることが示された。それらのパラメータはそれぞれ1次元のスカラ量で表現されるため、単独では学習が比較的容易であると考えられる。しかし、旋法の場合は多くの音の組み合わせであるため、より難易度の高い課題であると考えられる。強化学習の音楽への応用については、感情に関する研究以外にも、音楽スタイルを楽譜から学習し、その曲のスタイルを模倣した音楽を生成する研究 [29] などがある。強化学習を用いて旋法を感情に適応させる研究は、私の知る限りでは本研究が最初である。

3.3 強化学習

3.3.1 基本的な枠組み

本節では、旋法の獲得に用いる強化学習の数理的枠組みについて述べる。強化学習は教師無し機械学習法の一つであり、ある環境内に置かれたエージェントに、試行錯誤に基づく経験から適切な行動を学習させる方法論である。強化学習の枠組みでは、環境は状態空間 S 内をタイムステップ t ごとに遷移を繰り返す。エージェントは、状態 $s \in S$ を観測し、遷移ごとにある行動 $a \in A(s)$ ($a \in A(s)$ は状態 s でとることができる行動の集合) をとって環境に働きかけ、行動の結果の善し悪しに応じて環境から報酬 r を受け取る。エージェントは次式などの形で表される r の関数としての収益 R_t を最大化できるように学習を行う:

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}, \quad (3.1)$$

ここで、 γ は $0 \leq \gamma \leq 1$ の割引率と呼ばれるパラメータである。 γ が 0 のときはエージェントは即時報酬 r_{t+1} のみを考慮して学習するが、 γ が 1 に近づくにつれ、は未来の報酬をより多く考慮に入れるようになる。エージェントは r をもとに、状態 s の価値 $V(s)$ や、 s における行動 a の価値 $Q(s, a)$ を推定してゆく。 $V(s)$ および $Q(s, a)$ は次式で定義される:

$$V(s) = E\{R_t | s_t = s\}, \quad (3.2)$$

$$Q(s, a) = E\{R_t | s_t = s, a_t = a\}. \quad (3.3)$$

ここで、 γ は割引率と呼ばれる 0 から 1 までの値であり、収益 R_t は報酬 r の系列の関数として定義される。 γ が大きくなるにつれ、エージェントは長期的な展望を重視するようになる。

強化学習において、多くの場合、状態は有限かつ離散的であり、環境は基本的に MDP と呼ばれる形式をとることが仮定される。MDP においては、状態遷移は現在の状態にのみ依存して遷移先が決まる一次のマルコフ過程であり、エージェントが行動 a をとったとき、環境は状態が s であれば状態 s' に確率 $P_{ss'}^a$ で遷移する。また、エージェントの行動 a に対して環境は、期待値 $R_{ss'}^a$ の報酬 r を与える。このように MDP は $P_{ss'}^a$ と $R_{ss'}^a$ で特徴づけられる。

他方、エージェントは方策 (policy) と呼ばれる行動指針 π をもち、状態 s において、ある確率 $\pi(s, a)$ で行動 a をとる。 $\pi(s, a)$ の値は $V(s)$ や $Q(s, a)$ を参照しながら決定することが多く、 $V(s)$ や $Q(s, a)$ は $\pi(s, a)$ と共に改善されていくことが期待される。

もし環境の $P_{ss'}^a$ と $R_{ss'}^a$ が既知の定数でかつ $\gamma < 1$ あれば、 $V(s)$ と $Q(s, a)$ は Bellman 方程式と呼ばれる連立方程式を用いて解くことができる。しかし多くの場合、未知環境の下での学習が問題となる。未知環境の下では、 $\pi(s, a)$ の下での試行錯誤から経験的に学習を行う。

3.3.2 ブートストラッピングとエピソード

未知環境での強化学習には様々なタイプの手法がある。中でも、逐次更新型とエピソード単位の一括更新型の違いは本研究にとって重要であるため、以下に解説する。

一般に価値関数の改善は、次式のような $V(s)$ や $Q(s, a)$ を目標値に近づける更新によって行われる:

$$V(s_t) \leftarrow V(s_t) + \alpha[R_t - V(s_t)], \quad (3.4)$$

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[R_t - Q(s_t, a_t)], \quad (3.5)$$

ここで α はワンステップの学習の度合いを表すパラメータである。式 (3.1) で表される収益は、収益が実際に計算できるようになるまでの時間が無限大であるため、厳密には計算ができない。そこで、

$$\begin{aligned} V(s) &= E\{R_t | s_t = s\} \\ &= E\left\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s\right\} \\ &= E\left\{r_{t+1} + \gamma \sum_{k=0}^{\infty} \gamma^k r_{t+k+2} | s_t = s\right\} \\ &= E\{r_{t+1} + \gamma V(s_{t+1}) | s_t = s\} \end{aligned} \quad (3.6)$$

となることを用い、 R_t を $(r_{t+1} + \gamma V(s_{t+1}))$ を用いて近似することで、逐次的な更新を行う手法がある。この手法は TD(0) と呼ばれる手法であり、 $V(s_t)$ を他の価値関数の推定値を用いて間接的に推定する仕組みはブートストラッピングと呼ばれる。この近似により、TD(0) の更新式は次のようになる:

$$V(s_t) \leftarrow V(s_t) + \alpha[r_{t+1} + \gamma V(s_{t+1}) - V(s_t)]. \quad (3.7)$$

しかし、 R_t を直接計算できる場合もある。経験する状態の遷移が有限の長さで終了するエピソードを単位とした反復である場合は、エピソードの終了を待てば R_t の値が確定する。この R_t を直接用いて、エピソードに出現した各状態の状態価値を式 (3.4) のように一括更新することができる。この手法はモンテカルロ法と呼ばれる。 Q の更新に関するにも同様の方法がとれる。

ブートストラッピングによる逐次更新型の手法には TD 学習、Q 学習、Sarsa などがあり、エピソード単位の手法には、モンテカルロ法、プロフィット・シェアリングなどがある。本研究ではモンテカルロ法を用いる。

3.3.3 関数近似、最急降下法

強化学習を行うにあたっては、状態と行動の全ての組み合わせ (s, a) に対して $Q(s, a)$ の訓練を十分な回数行う必要がある。しかし、そのためには大きな時間ステップ数やデータ数が必要となる。そこで、経験した (s, a) から経験していない (s, a) へと経験を汎化することが重要となる。そこで、それを可能にする工夫として関数近似という手段を導入する。関数近似とは、 (s, a) の組み合わせの数よりも小さな要素数 n のパラメータベクトル $\vec{\theta} = (\theta_0, \theta_1, \dots, \theta_{n-1})^T$ の関数として $Q(s, a)$ を近似することである。それが可能であれば、より少ない経験から全ての $Q(s, a)$ の値を効率的に学習することができる。

関数近似の最もシンプル形態の一つは、関数をパラメータに対して線形に近似することである。線形近似においては $\vec{\theta}$ に対応する特徴ベクトル $\vec{\phi}(s, a) = (\phi_0(s, a), \phi_1(s, a), \dots, \phi_{n-1}(s, a))^T$ を用いて

$$Q(s, a) = \vec{\theta}^T \cdot \vec{\phi} = \sum_{i=0}^{n-1} \theta_i \cdot \phi_i(s, a). \quad (3.8)$$

のように $Q(s, a)$ を近似する。

$Q(s, a)$ が $\vec{\theta}$ を用いて汎化される場合、 $\vec{\theta}$ の更新においては最急降下法 (Gradient Descent) が多く用いられる。最急降下法とは、目的関数を最小化するために関数の勾配を計算し、最も急に関数が減少する方向にパラメータを更新するものである。本研究では、目的関数をターゲットとの二乗誤差とし、実例を得るたびにそれを最も減少させる方向にパラメータベクトルを修正する。二乗誤差の勾配は

$$\nabla_{\vec{\theta}} [R_t - Q(s, a)]^2 = -2[R_t - Q(s, a)] \nabla_{\vec{\theta}} Q(s, a). \quad (3.9)$$

であるから、これを用いてパラメータベクトルの更新式は

$$\vec{\theta} \leftarrow \vec{\theta} + \alpha [R_t - Q(s, a)] \nabla_{\vec{\theta}} Q(s, a). \quad (3.10)$$

と表すことができる。式 (3.8) で述べた線形近似を行う場合、勾配は

$$\begin{aligned}\nabla_{\vec{\theta}} Q(s, a) &= \left(\frac{\partial Q(s, a)}{\partial \theta_0}, \dots, \frac{\partial Q(s, a)}{\partial \theta_{n-1}} \right)^T \\ &= (\phi_0(s, a), \phi_1(s, a), \dots, \phi_{n-1}(s, a))^T \\ &= \vec{\phi}(s, a),\end{aligned}\tag{3.11}$$

のように簡単に計算でき、更新式は

$$\vec{\theta} \leftarrow \vec{\theta} + \alpha [R_t - Q(s, a)] \cdot \vec{\phi}(s, a).\tag{3.12}$$

となる。

3.4 提案手法

3.4.1 目的

本研究の目的はある感情 E に対して「 E を表現する旋法」を得ることである。そこで、まずこの概念を定義しておく。ただし簡単のため調律は 12 平均律とする²。

定義 1 12 平均率のもとでの 12 のピッチクラスの集合 $\mathbb{Z}_{12} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$ を状態空間 S として設定する。ここで S 上の 1 次のマルコフ連鎖 M および M から生成された時系列としての音列 $e = [s_1, s_2, \dots, s_T]$ を考える。 e を評価者が聴取して、平均的に E を良く表していると評価されるマルコフ連鎖 M_E が得られたとする。そのときこの M_E を「 E を表現する旋法」と呼ぶこととする。

この定義によって、本研究では旋法を、単なるピッチクラスの出現確率としての音階と区別し、ピッチクラス間の遷移傾向を含めたものとして扱う。

3.4.2 モンテカルロ法

E を表現する旋法 M_E を獲得するため、モンテカルロ法を応用した手法を以下に提案する。行動 $a \in A(s)$ は前のピッチクラス s から次のピッチクラス s' への移動音程を表すものとする ($\forall s \in S, A(s) = \mathbb{Z}_{12}$)。 s' は s と a によって一意に定まり ($P_{ss'}^a = 1$)、 $s' = (s + a) \bmod 12$ である。旋法の学習には、エピソードとして音列 $e = [s_1, s_2, \dots, s_T]$ を用いる。この音列は後に定義する $\pi(s, a)$ によって確率的に生成される。このエピソードが終わった時点で評価者の人間が評価を行う。その評価値を報酬 r_T とする。ここで、この報酬は e 全体に対する評価であるため、 γ を 1 として割引を行わない。この r_T をもとに、エピソード内に出現した各 s について式 (3.4) および式 (3.5) で価値関数を更新する。

² 本章の定式化は一般の 1 オクターブ n 分割 ($n > 0$) の調律システムに容易に拡張できる。ただし音程の扱いの上では、音程のシフトに対して音程幅が厳密に一定である n 平均律が望ましい。

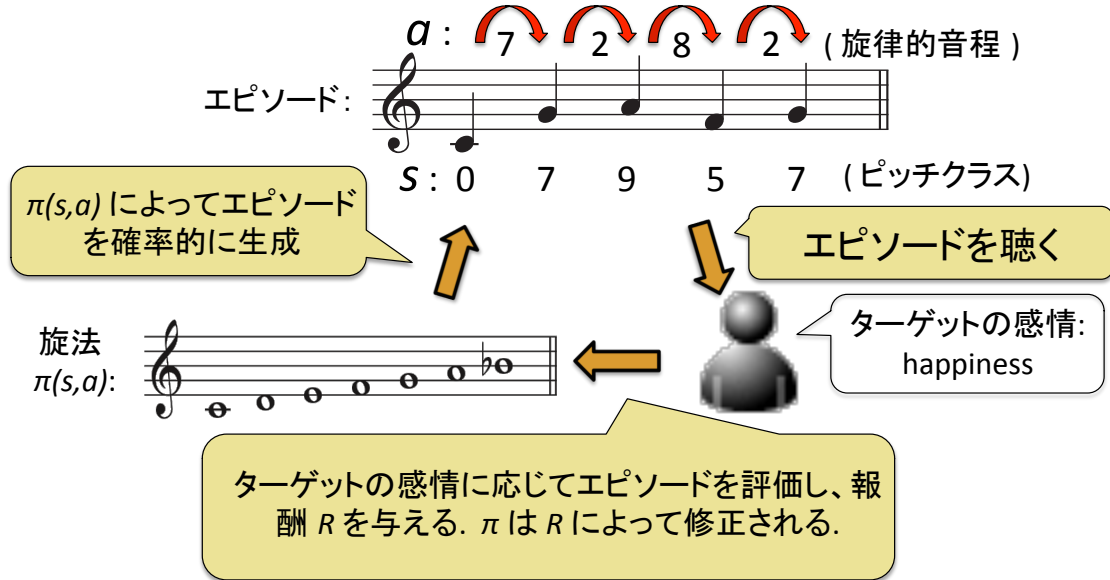


図 3.1. システムと評価者によるインタラクティブな強化学習の学習のサイクル. π が収束したと見なされるまでこのサイクルを繰り返す。

r_T は、エピソードの音列全体がどれだけ感情 E に対する表現力があるか、という $\{-3, -2, -1, 0, 1, 2, 3\}$ の 7 段階の整数値として設定する。各値の意味は、3: とても高い, 2: 高い, 1: やや高い, 0: どちらでもない, -1: やや低い, -2: 低い, -3: とても低い, とする。

このような報酬のもとで、価値関数 $V(s)$ は、感情 E を表現するためのピッチクラス s 必要性の度合いを、行動価値関数 $Q(s, a)$ はあるピッチクラス s から音程を a だけ移動し、 $s' = (s + a) \bmod 12$ へ移動することの必要性の度合いを表しているものだと考えられる。

ここで、方策 π の意味について言及しておく、 $\pi(s, a)$ は s から s' への遷移確率 $Pr(s'|s)$ に等しい。したがって、 π はマルコフ連鎖を導く。感情 E の学習が収束した際 π は先に定義した、 E を表現する旋法 M_E に一致する。

図 3.1 にこのような学習サイクルの概念図を示す。

3.4.3 方策

エピソードの生成に用いる方策 $\pi(s, a)$ は次のように定義する：

$$\pi(s, a) = \begin{cases} \frac{Q(s, a)}{\sum_{u \text{ s.t. } 0 < Q(s, u)} Q(s, u)} & \text{if } 0 < Q(s, a) \\ 0 & \text{otherwise.} \end{cases} \quad (3.13)$$

このように方策 π を定義することで、 s においてプラスの行動価値関数 $Q(s, a)$ をもつ行動 a は、 $Q(s, a)$ に比例する確率で出現し、マイナスの行動価値をもつ a は全く出現しなくなる。それにより、目標となる感情を表現するのに必要のない行動を排除することができると考えられる。

ただし注意すべきことは、一度 $Q(s, a) < 0$ となると、その a は s において全く出現しなくなるため、必要な音が運悪く偶然的に $Q(s, a) < 0$ となることを防ぐ必要があるということである。それを防ぐため、 $Q(s, a)$ の初期値を全て報酬の最大値である 3.0 として大きく見積もった初期値 (optimistic initial value) に設定する。このように大きく見積もられた初期値設定は、状態行動空間の広い探索を促す効果もあることが知られている。

この手法での学習が収束し、評価者によって高評価を得られるようになったとき、 π によって得られるマルコフ連鎖は、先に定義した「E を表現する旋法」となる。

3.4.4 4つの関数近似モデル

$Q(s, a)$ は、状態と行動の直積 $\mathbb{Z}_{12} \times \mathbb{Z}_{12}$ を定義域とする関数であり、 $12 \times 12 = 144$ 個の値をもつ。これら全てを愚直に学習するのは非効率である。そのため、3.3.3 節で述べた関数近似によって、比較のため4通りの近似モデルを構築する。

これらの近似においては、ピッチクラス s とは無関係に旋律的音程 a の価値が定まるという仮定を基本とし、4つのモデルに共通するパラメータ $\vec{\theta}$ を設定する。 $\vec{\theta}$ の要素数は 24 とし、 θ_0 から θ_{11} をピッチクラス 0 から 11 の状態の価値に相当するパラメータ、 θ_{12} から θ_{23} は、ピッチクラスの遷移 0 から 11 の行動の価値に相当するパラメータとして用いる。

モデル 1: 音高モデル

最初のモデルは、 a による旋律的音程の遷移の価値を無視し、遷移先のピッチクラス $s' = (s + a) \bmod 12$ の価値のみから $Q(s, a)$ を近似するモデルである。このモデルは式 (3.8) の線形近似において特徴ベクトルは次のようにとったものに相当する:

$$\phi_i(s, a) = \phi_i(s') = \begin{cases} 1 & \text{if } i = s' \\ 0 & \text{otherwise.} \end{cases} \quad (3.14)$$

このモデルでの行動価値関数 $Q_1(s, a)$ は

$$Q_1(s, a) = \sum_{i=0}^{23} \theta_i \cdot \phi_i(s, a) = \theta_{s'}. \quad (3.15)$$

となる。このとき式 (3.13) の π によるマルコフ連鎖においては $Pr(s'|s) = Pr(s')$ となり、 M は現在の状態 s に依存せず、単なる確率分布に退化する。

モデル 2: 音程モデル

2つ目のモデルは、状態 s に依存せず、 $Q(s, a)$ をピッチクラスの遷移 a の価値のみから近似するモデルである。このモデルは式 (3.8) の線形近似において特徴ベクトルを次のようにとったものに相当する:

$$\phi_i(s, a) = \phi_i(a) = \begin{cases} 1 & \text{if } i = 12 + a \\ 0 & \text{otherwise.} \end{cases} \quad (3.16)$$

このモデルでの行動価値関数 $Q_2(s, a)$ は

$$Q_2(s, a) = \sum_{i=0}^{23} \theta_i \cdot \phi_i(s, a) = \theta_{12+a} \quad (3.17)$$

となる。このとき式 (3.13) の π によるマルコフ連鎖においては $Pr(s'|s) = Pr(a)$ となり、 M は現在の状態 s に依存せず、 a に関する確率分布に退化する。

モデル 3: 加算モデル

3番目のモデルは、行動価値 $Q(s, a)$ を遷移先のピッチクラス s' およびピッチクラスの遷移 a の価値の両方を考慮し、その和で近似するモデルである。これは式 (3.8) の線形近似において特徴ベクトルを次のようにとったものに相当する:

$$\phi_i(s, a) = \begin{cases} 1 & \text{if } i = s' \text{ or } i = 12 + a \\ 0 & \text{otherwise.} \end{cases} \quad (3.18)$$

このモデルでの行動価値関数 $Q_3(s, a)$ は

$$Q_3(s, a) = \sum_{i=0}^{23} \theta_i \cdot \phi_i(s, a) = \theta_{s'} + \theta_{12+a} \quad (3.19)$$

となる。

モデル 4: シグモイド乗算モデル

4番目のモデルは、遷移先のピッチクラス s' およびピッチクラスの遷移 a の価値の貢献度が相乗効果をもつように設計されたものである。行動価値関数は、 s' と a のそれぞれに対応するパラメータをシグモイド関数 (図 3.2)

$$\sigma(\theta) = \frac{1}{(1 + e^{-\theta})}. \quad (3.20)$$

で変換したものの積を用いて近似される。

シグモイド関数は $\theta \rightarrow -\infty$ で 0, $\theta \rightarrow \infty$ で 1 に漸近するため、 $\sigma(\theta)$ の範囲をコントロールすることができる。

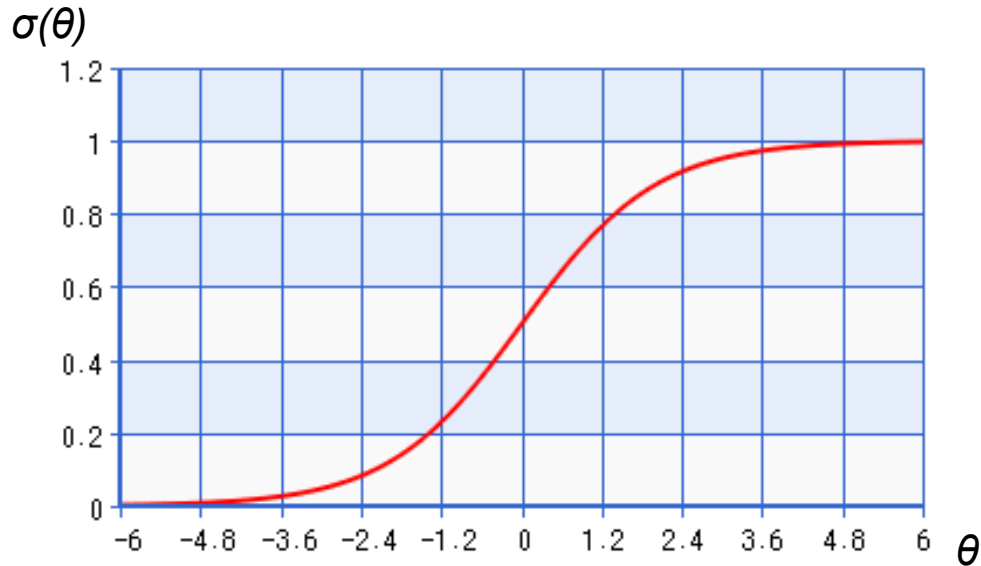


図 3.2. シグモイド関数.

これを用いて、このモデルでの行動価値関数 $Q_4(s, a)$ は

$$Q_4(s, a) = 6\sigma(\theta_{s'})\sigma(\theta_{12+a}) - 3. \quad (3.21)$$

と表される。ここで係数と切片は $-3 < Q_4(s, a) < 3$ となるように設定した。シグモイド関数の微分は

$$\frac{d\sigma(\theta)}{d\theta} = \sigma(\theta)(1 - \sigma(\theta)), \quad (3.22)$$

であるため、 $Q_4(s, a)$ の勾配は偏微分

$$\frac{\partial Q_4(s, a)}{\partial \theta_i} = \begin{cases} 6\sigma(\theta_{s'})(1 - \sigma(\theta_{s'}))\sigma(\theta_{12+a}) & (i = s') \\ 6\sigma(\theta_{12+a})(1 - \sigma(\theta_{12+a}))\sigma(\theta_{s'}) & (i = 12 + a) \\ 0 & (\textit{otherwise}). \end{cases} \quad (3.23)$$

を用いて計算できる。

3.5 実験

前節で述べた行動価値関数 $Q(s, a)$ の4つの関数近似モデルを用い、感情を表現する旋法の学習が可能かどうか、どのモデルが有効かを比較するため、著者（評価者1）による実験を行った。本節ではその結果を報告する。なお $V(s)$ については関数近似を用いずに通常通り式 (3.4) で更新して求める。 $V(s)$ は旋法におけるピッチクラス集合の構成をみるために用いる。

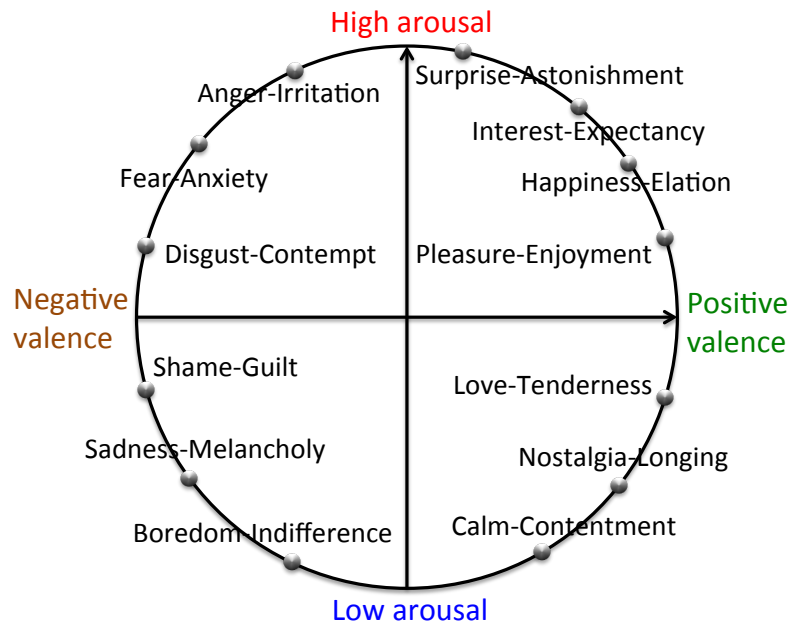


図 3.3. 感情カテゴリーの環状モデル. 左右の軸が快不快の度合い (Valence) を, 上下の軸が覚醒度の高低 (Arosal) を表し, 二次元の中に各感情カテゴリーが配置されるモデル.

3.5.1 実験条件

学習対象の感情の選択においては, Juslin らの環状モデル (circumplex model [30], 図 3.3) を参照し, 喜び (happiness), 悲しみ (sadness), 恐れ (fear), 優しさ (tenderness) の 4 つの感情を選択した. このモデルは, 快不快 (valence) と覚醒度 (arosal) の張る 2 次元平面上に, 諸感情を環状に配置したものである. 環状モデルの各象限から, 代表する感情のカテゴリーを 1 つずつ選択したものが上記の 4 つの感情である³.

エピソードとなる音列に関しては次のように設定した. 音列の長さは, 長すぎると学習が進まない恐れがあるため, 短く設定し 5 音とした. また, 旋法はトランスポーズしたものと同一の旋法になるため, 調の混同で学習が進まない恐れがある. そこで, 参照点となる主音を MIDI 番号 60 の C として設定し, エピソードの最初の音を C に固定した. また, 評価の精度を上げるため, 評価に確信が持てるまで音列を反復した. 12 個のピッチクラスは MIDI 番号 60 から 71 までの 12 音とした. 発音タイミングは 300ms ごとの等間隔であり, 音色は MIDI のピアノ音とした. ステップサイズパラメータ α は 0.1 に設定した.

³ ただし, 右下の象限について nostalgia ではなく tenderness を選択したのは, nostalgia がネガティブな感情も含まれる場合があると考えたためであり, ポジティブな感情とネガティブな感情のバランスを考慮し, ポジティブな語として tenderness を選択した.

3.5.2 実験結果

報酬の推移

図 3.4 にそれぞれの感情に関する 4 つのモデルの学習過程における、直近 10 エピソードの平均報酬を示す。図 3.4 から次のようなことがわかる。まず、いずれのモデルにおいても報酬の改善がみられ、報酬の期待値がおおむね 0 以上へ収束する。ただし、音程モデルにおいては得られる報酬がおおむね 0 から 1 の間と少ない。このことから、旋法においては、ピッチクラスの方が旋律的音程よりも重要な役割を果たしていると考えられる。

ピッチクラスモデルとシグモイド乗算モデルがいずれの感情においても比較的良い性能を示している。「優しさ」においてはシグモイド乗算モデルがピッチクラスモデルを上回っており、「恐れ」以外では 2 程度の「高い」報酬を、「恐れ」において 1 程度の「やや高い」報酬を平均的に獲得している。このことから、旋律的音程はピッチクラスよりも効果が小さいとはいえ、両方を用いた方がより感情を表現する力が強まると考えられる。ただし収束の早さに関しては、ピッチクラスモデルの方が早い。これは、音程に関するパラメータがない分パラメータ数が少ないことが関係しているものと推測される。

獲得した旋法

図 3.5 に学習後の $\vec{\theta}$ を図示する。

青のシグモイド乗算モデルは、ピッチクラスの価値に関わる θ_0 から θ_{11} までの起伏の特徴をピッチクラスモデルと共有し、旋律的音程の価値に関わる θ_{12} から θ_{23} までの起伏の特徴を音程モデルと共有しているさまが見てとれる。シグモイド乗算モデル以外においては勾配の各要素が 0 または 1 をとるのに対し、シグモイド乗算モデルの勾配においては状態と行動に相互依存性がある。シグモイド乗算モデルにおいては、この相互依存性が上手く働いたため、うまく適応できたという可能性がある。また、 θ_0 から θ_{11} の方が θ_{12} から θ_{23} よりも起伏がはっきりしていることから、ピッチクラスの方が音程の推移よりも感情の学習にとって重要であることが示唆される。次に、各モデルの状態価値 $V(s)$ を図 3.6 に示す。明確に $V(s) > 0$ であるような s の集合は、旋法において必要なピッチの要素のみを抽出した「音階」を表すものとみなせる。図 3.7 に音符の形での音階を示す。

最も高性能を示しているシグモイド乗算モデルを中心に $V(s)$ から読みとられるピッチクラスの構成を観察すると、音階としては喜びの旋法はミクソリディア旋法と同一である。悲しみの旋法はロクリア旋法に D と G を付加したものであり、恐れは他の旋法ほど明確ではないが悲しみの旋法に似ており、悲しみの旋法にさらに E および B を付加した旋法と捉えることもできる。優しさの旋法は長旋法（イオニア旋法）から F を除いた部分集合に一致する。全体としては否定的感情の方が構成音が多い傾向が読み取れる。

文献 [24] においては、各旋法に変換された音楽の抜粋をテンポと旋法の組み合わせを被験者に聴かせ、快不快 (Valence) と覚醒度 (Arosal) の張る二次元平面にマッピングする実験を行っている。テンポによる違いがあるため解釈が難しいが、少なくともその実験結果からは、

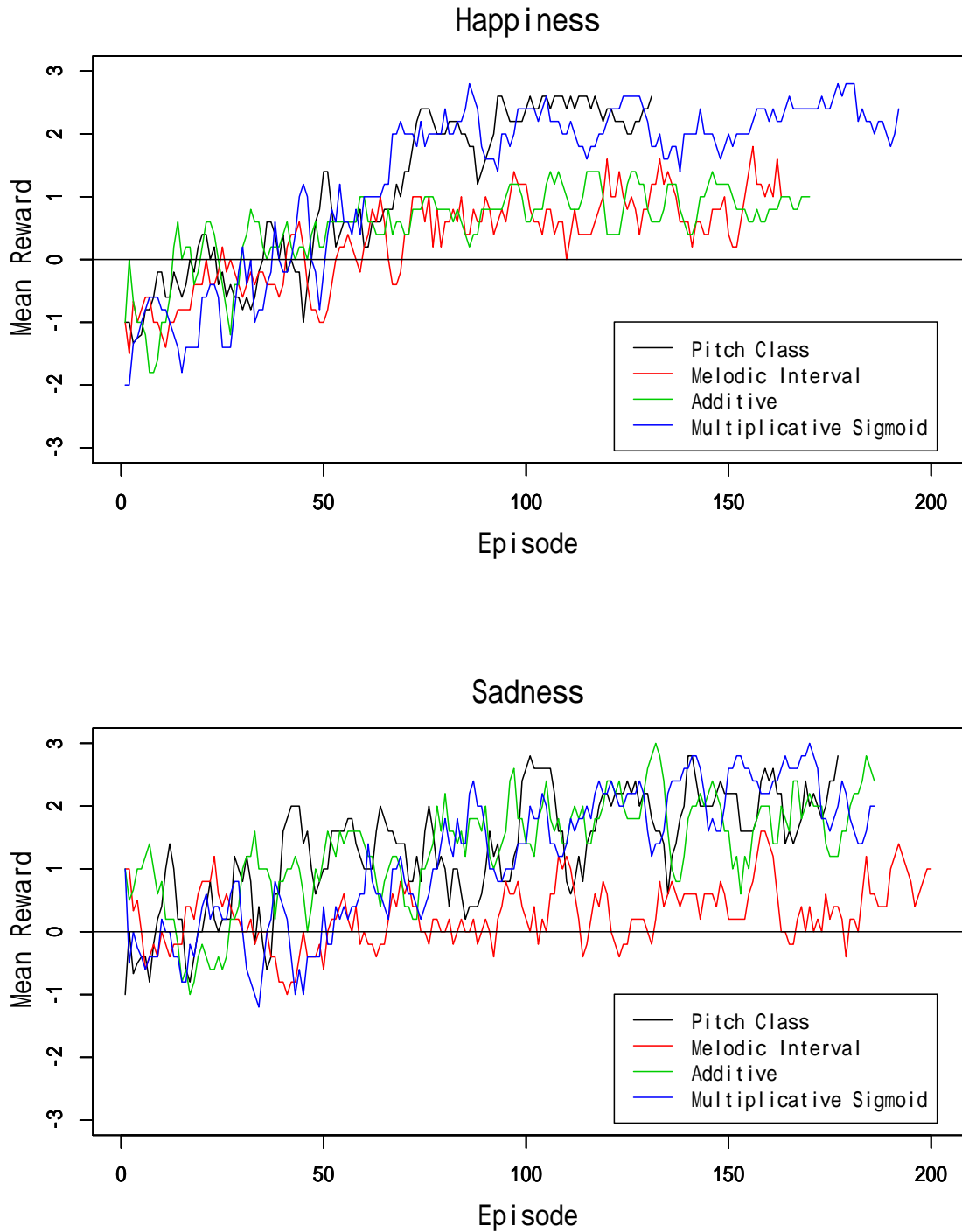


図 3.4. 平均報酬のエピソードごとの推移。次ページへ続く。

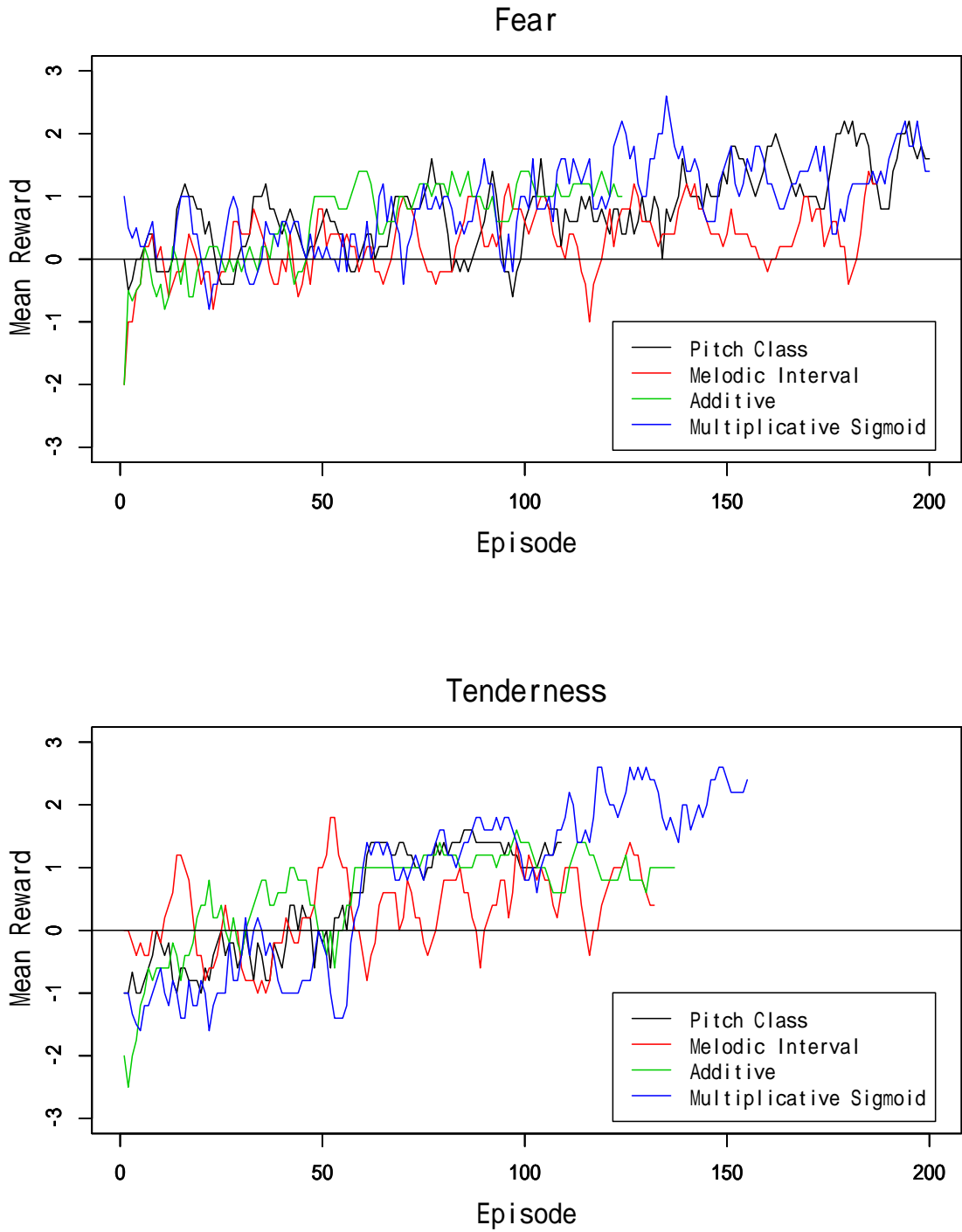


図 3.4. 平均報酬のエピソードごとの推移.

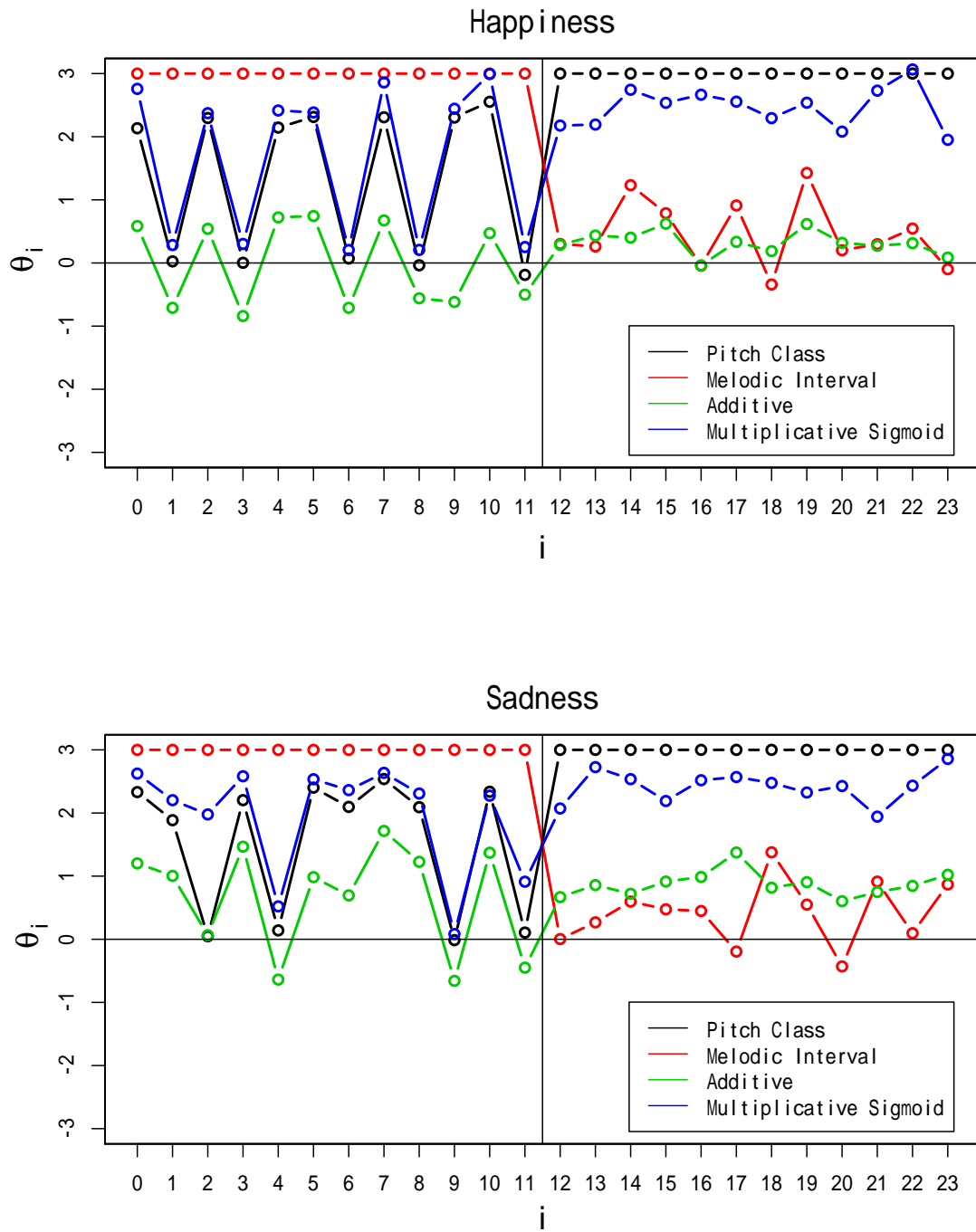


図 3.5. 学習後の $\vec{\theta}$. $i = 11$ までがピッチクラスの値に対応し, $i = 12$ 以降が旋律的音程の値に対応する. 次ページへ続く.

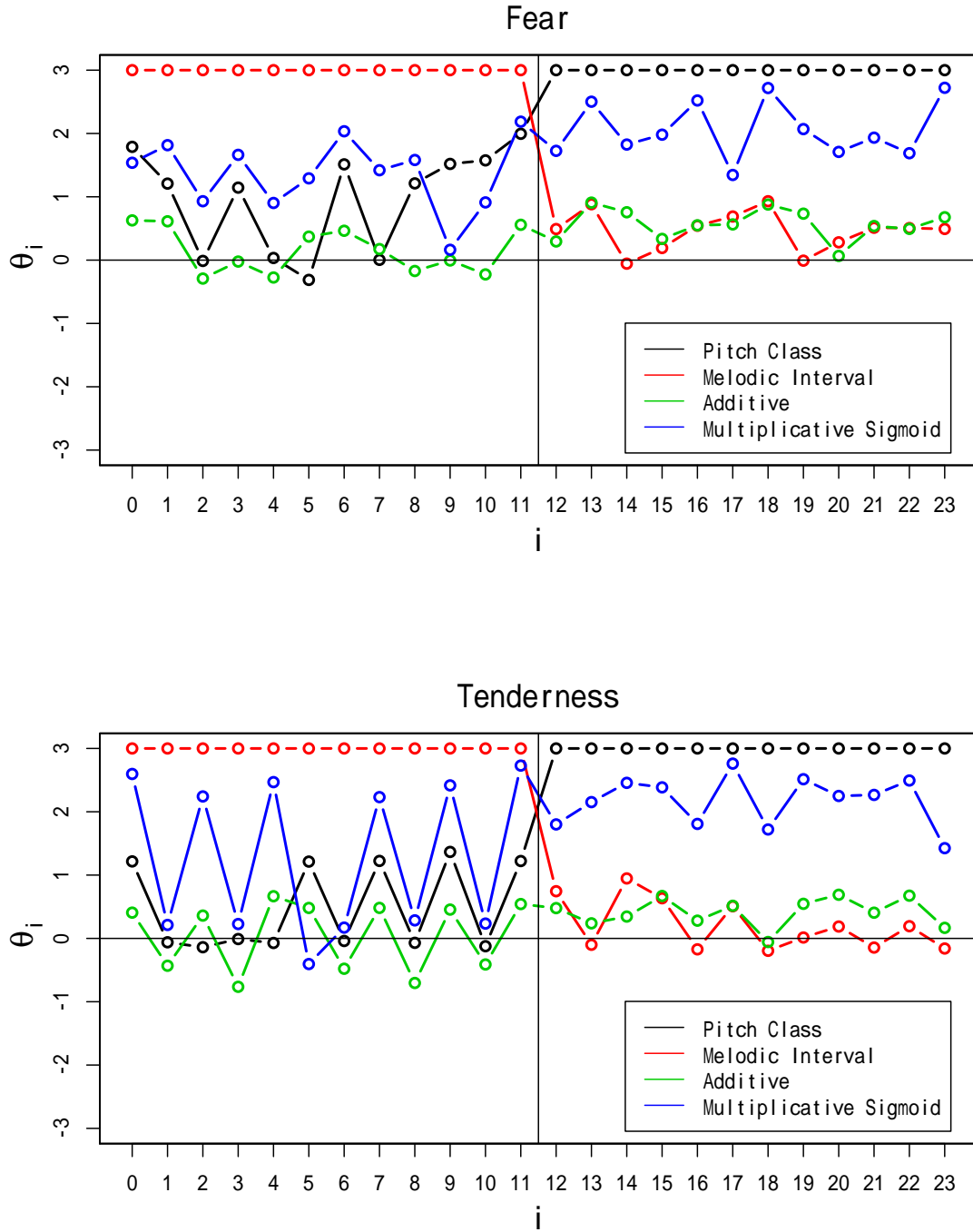


図 3.5. 学習後の $\vec{\theta}$. $i = 11$ までがピッチクラスの値に対応し, $i = 12$ 以降が旋律的音程の値に対応する.

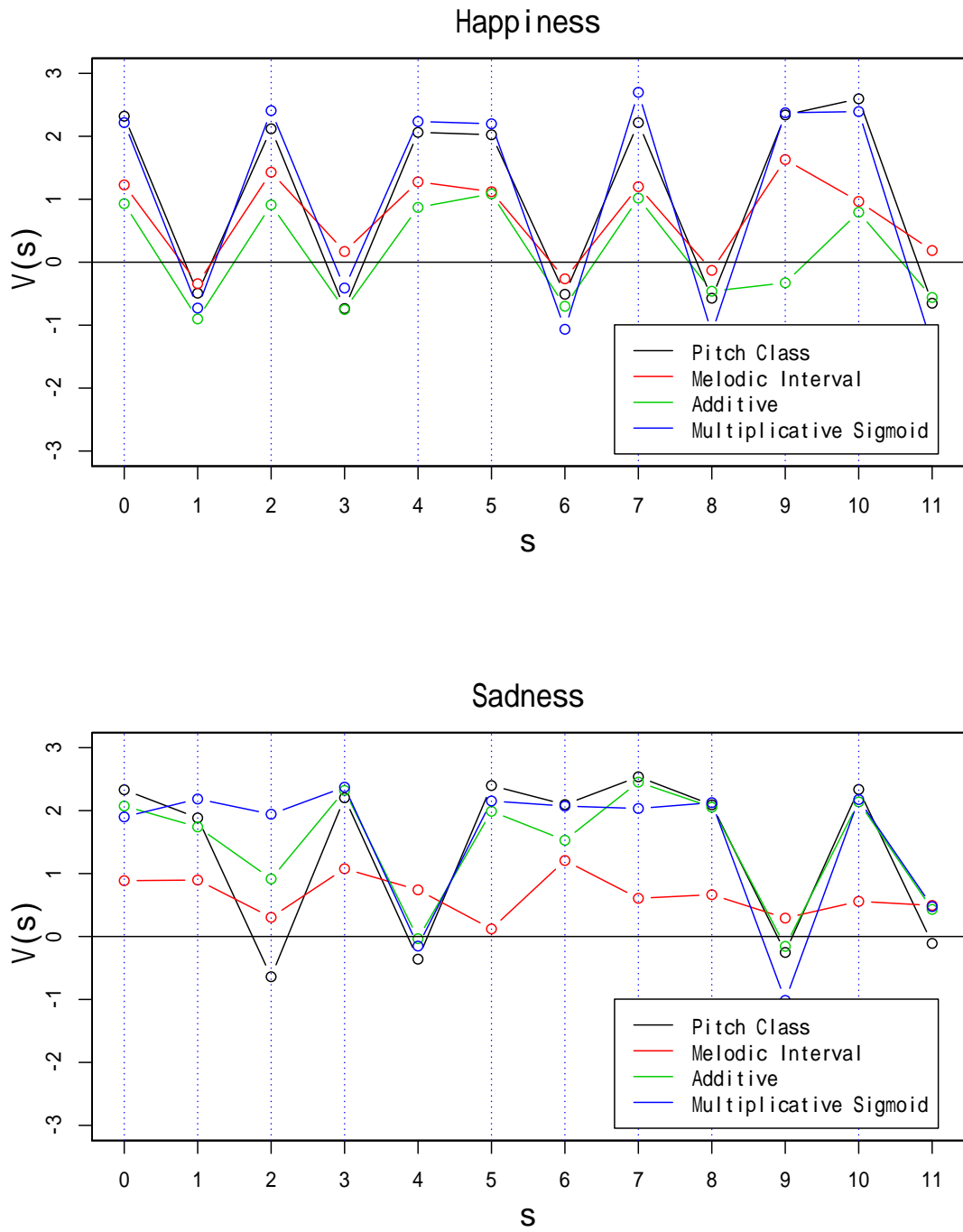


図 3.6. 学習後の $V(s)$. 次ページへ続く.

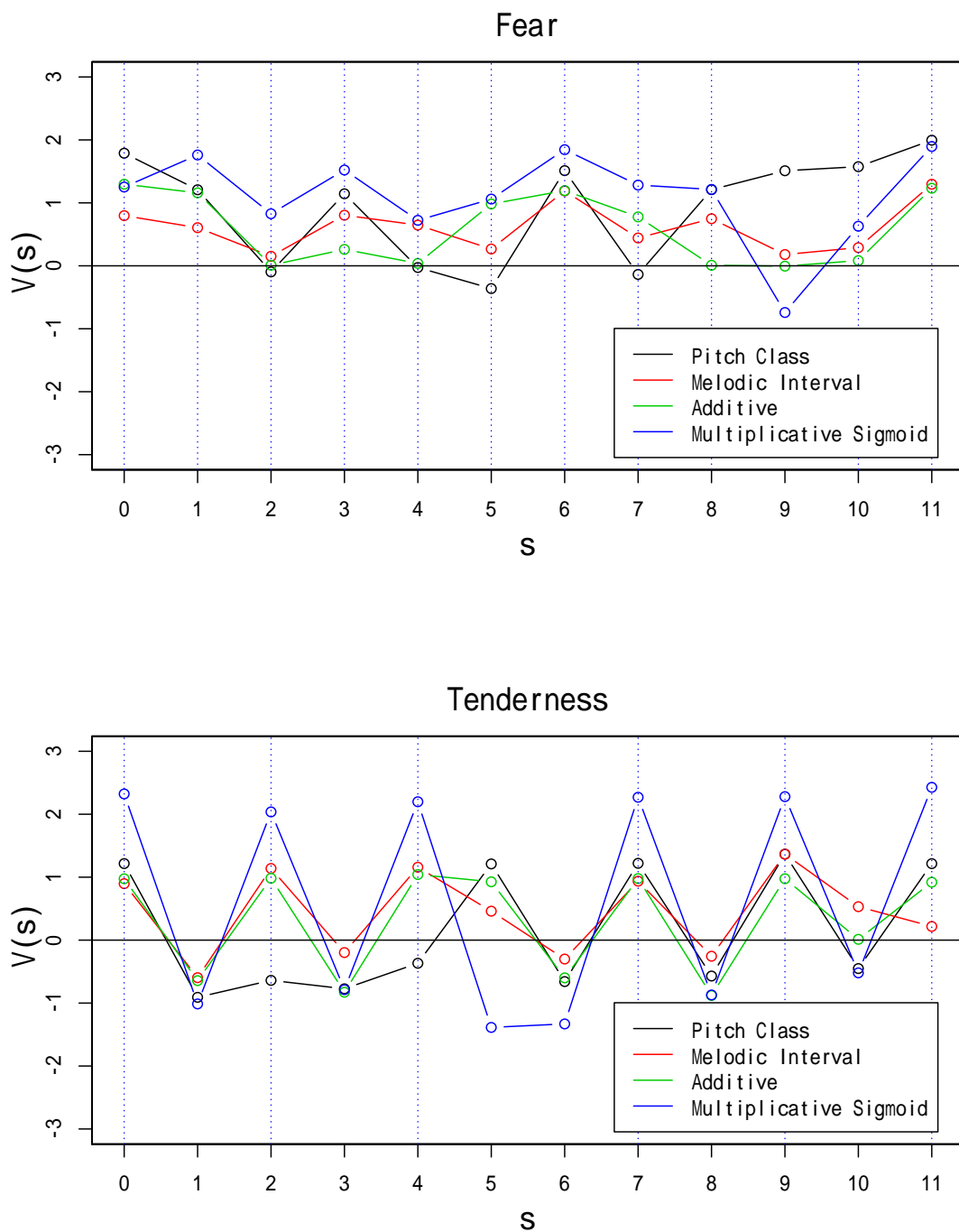


図 3.6. 学習後の $V(s)$.

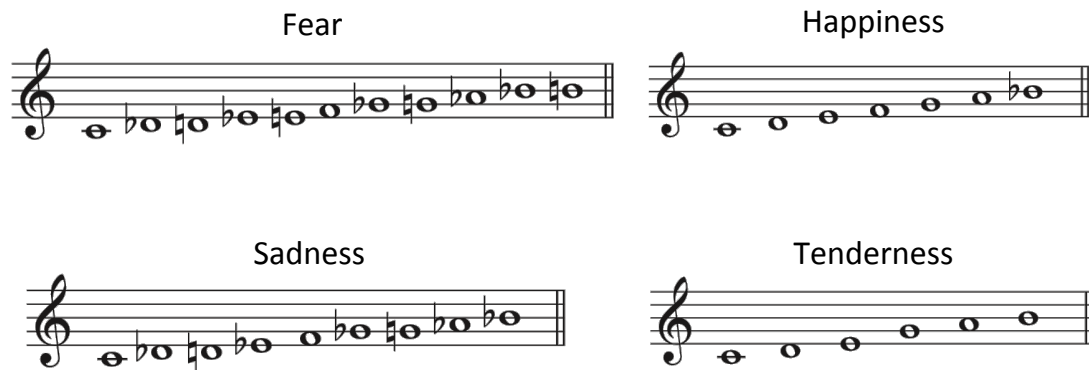


図 3.7. 学習後に得られた音階 ($V(s) > 0$ であることが明確な s の集合).

ミクソリディア旋法およびイオニア旋法は覚醒度が比較的高く、ロクリア旋法は覚醒度が比較的低い旋法であると解釈できる。このことは、本実験で覚醒度の高い「喜び」に対して学習した結果がミクソリディア旋法であったこと、覚醒度の低い「悲しみ」に対する学習の結果がロクリア旋法に近かったことと整合する。ただし、覚醒度の低い「優しさ」に対して学習した結果がイオニア旋法に近かったこととは整合しない。この理由を考えると、優しさの学習結果として得られたのが完全なイオニア旋法ではなく、一音欠けたイオニア旋法の部分集合であったことを考えれば、イオニア旋法の全てのピッチクラスは使わないということが、何らかの理由で覚醒度を相対的に下げることにつながり、優しさを表現できたのだという仮説を立てることができる。あるいは、シグモイド乗算モデルの構成音の 4 とピッチクラスモデルの構成音 5 が競合している可能性もある。これらは今後検証すべき課題である。

興味深いことに、日本の伝統的な旋法 [26] において、明るいことを意味する名の陽旋法 (図 3.8 左) および、暗いことを意味する名の陰旋法 (図 3.8 右) は、それぞれ喜びの旋法および悲しみの旋法の中に埋め込まれている。喜びの旋法は C 上のミクソリディア旋法と同一であり、長旋法とは B \flat のみが異なっているが、陽旋法はミクソリディア旋法の中に埋め込まれている。また、悲しみの旋法は自然的短音階に D \flat と G \flat を付加したものとも解釈できるが、悲しみの旋法の中には陰旋法が埋め込まれている。これらのことは、評価者が日本人であったことと関係し、文化と学習結果の間に整合性があることを意味するかもしれない。あるいは、これは文化の問題ではなく普遍性を意味する証拠である可能性も考えられる。それを明確にするにはより詳しい実験が必要であるが、本手法は、評価者に適応するという手法であるため、文化による差異の有無にかかわらず使用可能である可能性が高いと考えられる。

追加実験

本手法が他の評価者でも学習可能であることを確認するため、加算モデル以外の 3 種類のモデルと 4 種類の感情の 12 通りに対して他の 1 名の評価者 (評価者 2) によって同様の学習を行った。その際の平均報酬の推移を図 3.9 に掲載する。この評価者においても、平均報酬は右肩上がりの傾向を示しており、学習が可能であることがわかる。ただし、モデル間の性能の

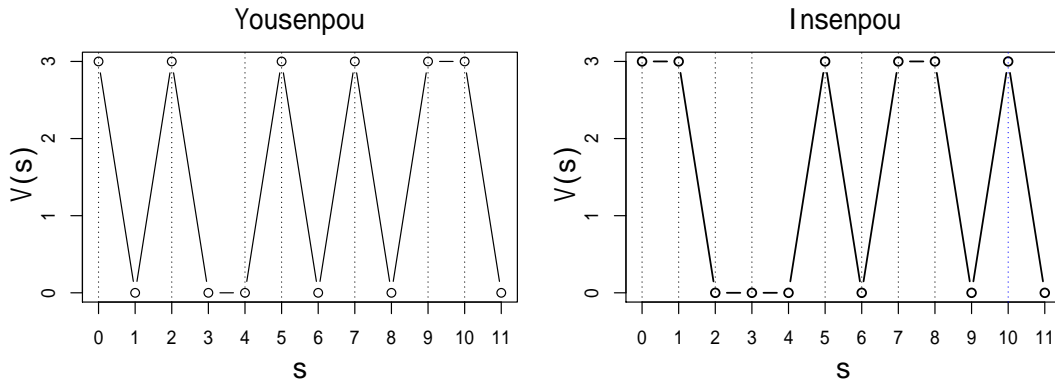


図 3.8. 陽旋法と陰旋法（実線部分）[26]. ただし上行形と下降形の両者を混合したもの. 陽旋法は喜びの旋法（点線部分）に、陰旋法は悲しみの旋法（点線部分）に埋め込まれている.

差の優劣はわかりにくく、定量的な評価を行うためにはより多くの評価者によるデータが必要となるだろう.

3.6 第2のモデルについての考察

前節の実験においては、第2の関数近似モデル（音程モデル）はピッチを考慮した他のモデルに比べて感情を表現する力は劣るという結果であったものの、このモデルには音楽理論的に興味深い点がある. 本節ではその点について指摘し、このモデルから「音程スケール」という新しい概念を導く.

通常、音階（scale）という概念はピッチクラス集合の構成を表し、どのような音程がピッチクラス間に生じるかについては明示化しないことが多い. しかし、それを逆にした概念を考えてみる. すなわち音程の集合としての「スケール」であり、どのようなピッチクラス集合が出現するかを直接明示しないような概念である. これはまさに第2のモデルが学習の対象としていたものである. 私はこれを通常の音階（スケール）と対比して「音程スケール」と名づける. そして本章では、通常の音階の方をこの概念と対比させて「音高スケール」とも呼ぶ.

ここで、さらなる音程スケールを特徴付けとして、本研究の第1と第2のモデルとの関連した次のような対概念を定義する.

定義 2 第1のモデルによる感情 E の強化学習が収束し、0以上の平均値を得られるようになったとき、 $\{s \in \mathbb{Z}_{12} | \theta_s > 0\}$ を、感情 E を表現する音高スケールと呼ぶ.

定義 3 第2のモデルによる感情 E の強化学習が収束し、0以上の平均値を得られるようになったとき、 $\{a \in \mathbb{Z}_{12} | \theta_{12+a} > 0\}$ を、感情 E を表現する音程スケールと呼ぶ.

この定義と実験結果により、例えばミクソリディア旋法は喜びを表現する音高スケールであ

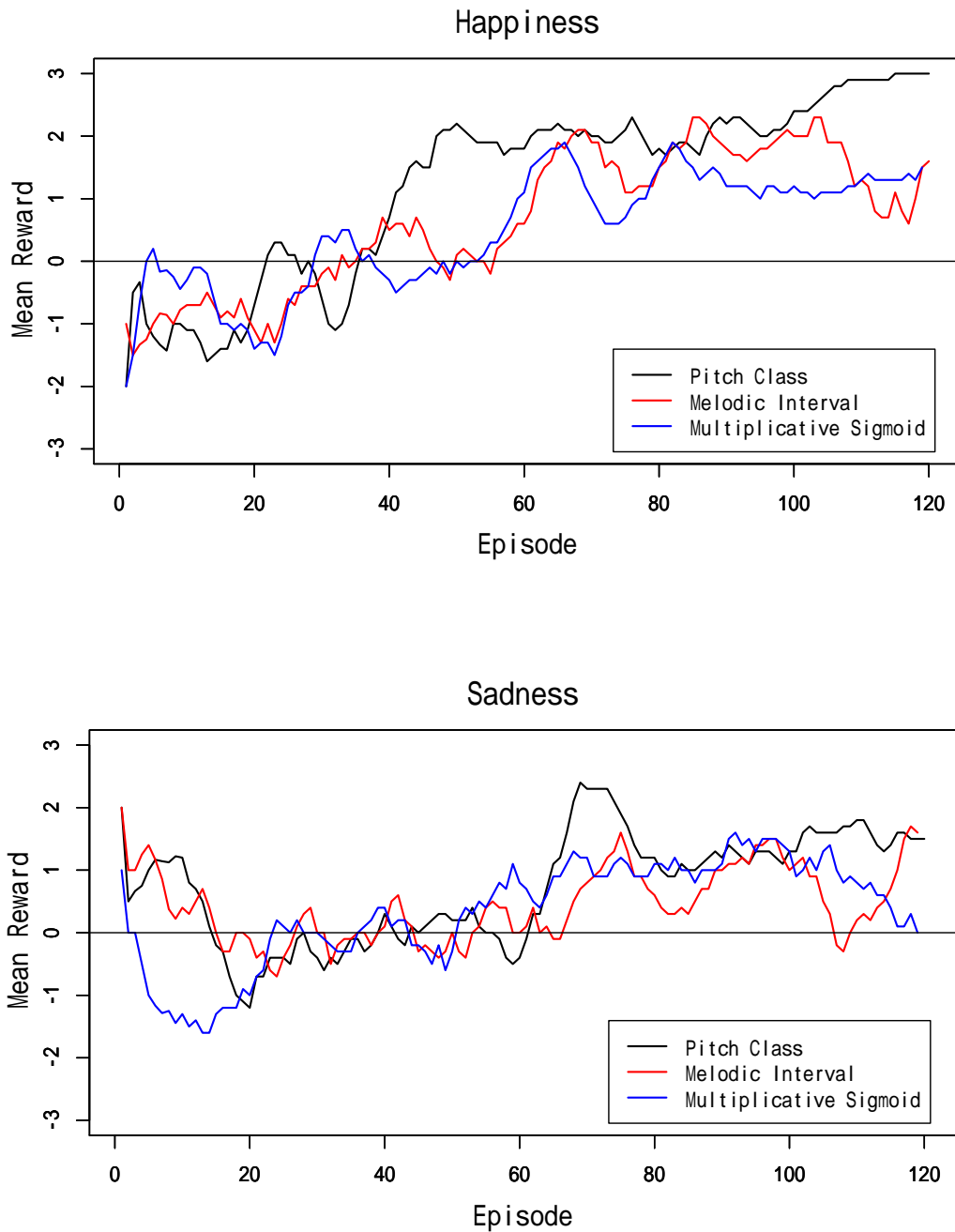


図 3.9. 平均報酬のエピソードごとの推移 (評価者2). 次ページへ続く.

り, $\{0,2,3,5,7,10\}$ は喜びを表現する音程スケールである. 前者に基づく楽曲を既存の楽曲から探すのは簡単であるが, 音程スケールやある感情を表現する音程スケールに基づいた楽曲は存在するかどうかはまだわからず, 現時点では理論的に考え出されただけのものにすぎない. 特に, 特定の感情を表現する音程スケールに基づく楽曲を探すのは, 単に音程スケールに基づ

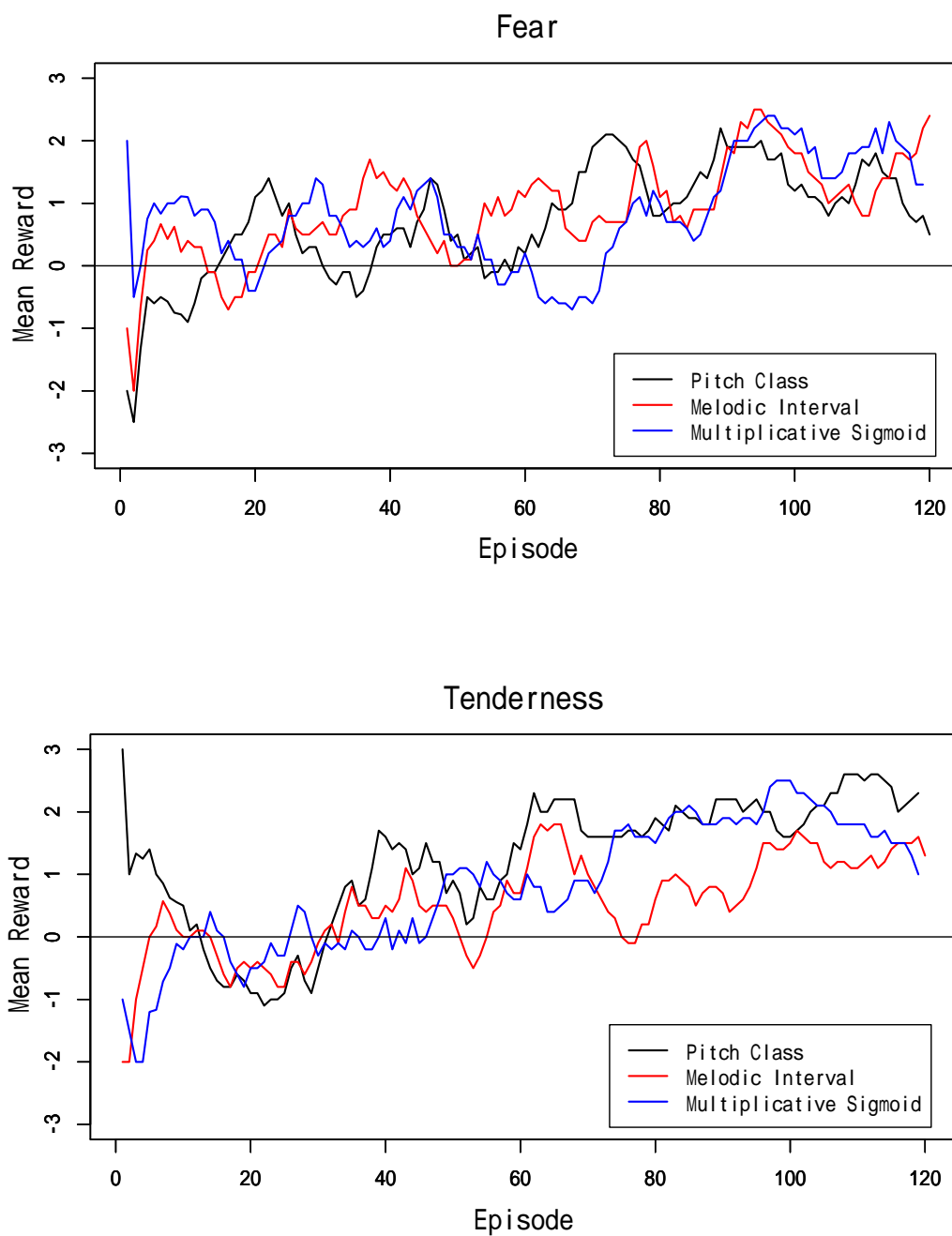


図 3.9. 平均報酬のエピソードごとの推移 (評価者 2).

く曲を探すよりさらに困難であろう。しかし今後、本手法を用いて得られる、感情を表現する音程スケールを作曲に応用することで、新しくそのような楽曲を作ることは可能性としてありうる。それによってこの概念の有用性が検証される可能性があるだろう。それは今後の課題として残される。

ところで、強化学習においては、行動を度外視した状態価値関数 $V(s)$ と、状態と行動の両者を考慮した行動価値関数 $Q(s,a)$ が用いられるが、状態を度外視した行動のみの価値関数

$$W(a) = E\{R_t | a_t = a\}. \quad (3.24)$$

が用いられないのはなぜだろうか？

これは、強化学習が取り扱う問題のほとんどが、ある状態に至ることが目的として設定され、行動は目的の状態へ至るための手段として位置づけられるからだと考えられる。しかし、音楽のように「動くことそのもの」にも目的がある場合、 $W(a)$ のようなものが重要になっても良いはずである。音程スケールはまさにそのような価値を担うべきものであり、検証する価値がある概念だろう。次章ではこの概念についての詳しい考察を行う。

3.7 章のまとめ

感情を表現する旋法を獲得するため、強化学習を用いた4つのモデルを提案し、それらの性能を比較した。悲しみ、喜び、恐れ、優しさの4つの感情について学習を行ったところ、最良のモデルであったシグモイド乗算モデルにおいては、喜び、悲しみ、優しさに対して表現力の高い旋法を、恐れと怒りに関してはやや高い表現力をもつ旋法を獲得することができた。モデル間の比較からは、感情の表現においてピッチクラス集合の構成の方が旋律的音程よりも重要であることが示唆された。それぞれの感情に適応した旋法のピッチクラスセットの構成は、喜びはミクソリディア旋法に一致する7音の旋法、悲しみはロクリア旋法にDとGを付加した9音の旋法、恐れは悲しみの旋法にさらにEとBを付加した11音の旋法、優しさはイオニア旋法の部分集合としての6音の旋法という結果になり、否定的な感情の方が音数が多くなる傾向が推測される。喜びと悲しみの旋法は、ピッチクラス集合の構成としては、日本の伝統的な旋法である陽旋法と陰旋法がそれぞれ埋め込まれたような旋法となった。また、第2のモデルからは旋法から音高の情報を捨象した音程スケールの概念が導かれた。

本研究に関連する発展的な課題として次のようなものが挙げられる。

1. より多くの感情のカテゴリに対する本手法の有効性や、感情のタイプによる結果の違いの分析：例えば、人間の主体に喚起される感情と客体的に認知する感情、というような感情のタイプの分け方など、本手法が学習対象とする感情のタイプによってモデルの効力がどう変わるかを調べることは、旋法という音楽組織そのものの役割や表現力の可能性と限界を明らかにするのに役立つと考えられる。
2. 微分音による音律などの他の調律システムに対して本手法を適用すること：これによって、なぜ伝統的に12音の調律システムが広く使用されているのか、12音ではない他の調律システムでは感情の表現力が弱いのかなどといった疑問に対し、調律システムの音数を変えて強化学習を行ったときの報酬の大きさを調べることで明らかになる可能性がある。

54 第3章 感情を表現する旋法の生成

3. 個人差や文化の違いによる結果の違いの調査：適応的な本手法により，評価者となる個人にとって感情を感じさせる旋法を生成できるとしても，それがどのくらい他の人にとって普遍性をもつのか，そして，個人差や文化の差によって異なる適応結果が得られるのかを明らかにすることは，旋法の研究にとっても，音楽一般の問題としても，興味深い研究テーマの一つである。
4. 音域，音色，テンポ，リズム，多重音など，より多くの音楽的要素を反映するように手法を拡張すること：本研究では旋法を学習対象としたが，各音楽的要素に対して同様の学習モデルを構築することができれば，それぞれの音楽的要素がどのような表現力の可能性を担っているのかを明らかにすることができるだろう。また，各要素を同時的に学習できるようになれば，旋法だけでなく，感情を表現する音楽作品の自動生成にもつながるだろう。しかし，同時に学習すべき要素が多くなると，状態や行動の空間が肥大化して学習時間が増加するため，価値関数や行動価値関数の近似を行うなどの対処がますます重要になってくると考えられる。
5. 音程スケールのより詳細な探求と作曲への応用：音程スケールの意義を検証するためには，既存の楽曲の中で関連するような楽曲がないかどうか調査し，あるとすればそれを分析したり，実際に音程スケールを用いて作曲してみる必要があるだろう。次章では音程スケールについて詳しく扱う。

第 4 章

音程スケールと群の生成系

4.1 旋法の変種としての音程スケール

前章では、旋法概念を、使用される音高（スケール）のみの集合としての音階と区別して音程の遷移傾向を含むものとして扱い、旋法の強化学習における関数近似モデルを構築した。その中の第二のモデルは、旋法における音程の要素を重視し、音高の情報を捨象した、ある意味で極端なモデルである。私はそこから「音程スケール (interval scale)」の概念を導いた。音程スケールは、このように旋法概念の変種としてとらえられるが、理論的に考え出されたものであるため、実際の音楽において有用であるかどうかは検証してみなければわからない。

例えば、音程スケールとしての音程の集合のとりかたは音階と同様多数存在するが、音程スケールが実際の楽曲の中でどのように用いられうるか、どのような集合を選択すべきかという問題は、ほとんど未知の問題である。そこで本章では、音程スケールに基づく解釈できるような楽曲の実例を見いだして分析するとともに、音程スケールが代数学における群の概念と密接に関わることを示し、群の生成の観点からその一般的な形式的特徴を論じる、そこで得られる知見は、音程スケールを実際に応用する際の基本的な選択基準を提供するだろう。

4.2 音程スケールの定義

ここで、音程スケールの形式的な定義を与えておく。12 平均律のような 1 オクターブを 12 音に分割する調律システムにおいては、C, C#(D ♭), D, ..., B の 12 個のピッチクラスの集合は、整数 \mathbb{Z} の $\text{mod } 12$ による剰余類のなす集合 \mathbb{Z}_{12} と同一視できる。この調律システムの中の音階は、 \mathbb{Z}_{12} の部分集合として表すことができる。0 から 11 までの整数 m によって代表されるクラスを \bar{m} と表記することに約束すると、例えばダイアトニック・スケールは $\{\bar{0}, \bar{2}, \bar{4}, \bar{5}, \bar{7}, \bar{9}, \bar{11}\}$ であり、クロマティック・スケールは $\{\bar{0}, \bar{1}, \bar{2}, \bar{3}, \bar{4}, \bar{5}, \bar{6}, \bar{7}, \bar{8}, \bar{9}, \bar{10}, \bar{11}\}$ である。ただし、本章では C を主音としたイオニア音階 $\{\bar{0}, \bar{2}, \bar{4}, \bar{5}, \bar{7}, \bar{9}, \bar{11}\}$ と G から始まるミクソリディア音階 $\{\bar{7}, \bar{9}, \bar{11}, \bar{0}, \bar{2}, \bar{4}, \bar{5}\}$ のような差異は考えず、ピッチクラス集合として等しいものは同じ音階であると考え。

同様に1オクターブを n 音に分割する調律システムは \mathbb{Z}_n と同一視できる. そして \mathbb{Z}_n の中での音階は \mathbb{Z}_n の部分集合として定義できる. ただし空集合は音階に含まないものとする.

ここで, 調律システム \mathbb{Z}_n の中での音程について考えると, \mathbb{Z}_n の中でとりうる全ての音程 (ordered pitch-class intervals¹) の集合 $\{y - x(\text{mod } n) | x, y \in \mathbb{Z}\}$ (ただし, \mathbb{Z} はピッチ全体の集合を表す) を考えるとやはり \mathbb{Z}_n となる. この概念を音階とのアナロジーから音程スケールと名づけ, 次のように定義する:

定義 1 (音程スケール (interval scale)). 音程スケールとは, ある調律システムの音程の集合としての \mathbb{Z}_n の部分集合のことをいう. ただし空集合は含まない.

音程スケールの命名のについては, Messiaen の作品あるいは概念である「音家と強度のモード (Mode de valeurs et d'intensités)」[32, 33] の類推から「音程のモード (“mode of intervals”）」と呼ぶことも考えられが, 本論文では音階 (scale) とのアナロジーを重視して「音程スケール」と呼ぶこととする.

音階とのアナロジーで考えたとき, 音程スケールに期待されるのは, 長調と短調の違いや旋法の違いがもたらすような雰囲気や色彩の差異を, 別の方法で実現することができるのではないかということである. ただし音階と異なるのは, 全てのピッチクラスを使用する無調音楽においては, ピッチクラスを制限する音階の概念は有効性が低くなるが, 音程スケールは, 音階のように使用できるピッチを直接制限しないために無調音楽と相性が良いということである. 無調音楽においては, 調性音楽における調の違いのような広く共有された, 感覚の差異を支える理論的基盤は未だ存在しないと思われるが, 音程スケールがそのような機能を持ち得るのではないかという可能性は, 検討する価値があると私は考える.

4.3 音高と音程の双対性

通常音階 (スケール) は, 出現しうる音高とそうでない音高を定めものである. 音階の中では出現しうる二つの音高の間には音程が生じるが, それは音高を指定した結果として付随的に現れるものとしてとらえられる. ここで逆の考え方をとり, 出現しうる音程とそうでない音程を先に定め, 音高は音程の動きの結果として付随的に現れるものとみる見方をとることもできる. 音程スケールは後者の観点をとる概念であり, その意味で「音階」と双対となる概念である.

音高の \mathbb{Z}_{12} は絶対的な音の高さを, 音程の \mathbb{Z}_{12} は基準点からの相対的な音の高さを表すが, 音の高さという共通するものを扱っているため, 両者にはどちらが優先されるかというコンフリクトが存在すると考えられる. 例えば記譜法においては, 音高は直接音符で表記されるが, 音程は音符と音符の間から間接的に読み取られることになる. また, 音階や和音は通常, 音高の集合として指定される. 音階の各要素間には音程が生じるが, それがどのような音程かは通

¹ 二つのピッチクラス間の音程は, 正負の符号の違いを別のもつと見なして ordered pitch-class interval と呼ばれる [31]. しかし, 誤解がなければ単に「音程」と呼ぶことにする.

常明示化されず、音程はやはり二次的なものとして扱われる。

しかし逆に、音程が重視される場合もある。音階においては、絶対的な音高ではなく隣接する音程が全音か半音かという観点から音程列として見られる場合がある。また、数字付き低音における和音の扱いは、和音をピッチの集合としてとらえず、ベース音からの音程度数によって表記することで、和音の形と転回形の差異が表される。数字付き低音においては、調における機能からは解釈しにくい和音も、ベース音の動きとの関係で理解しやすくなることがある。ピッチクラスの内包する全ての音程を表す音程行ベクトル (interval vector) の概念 [65] においては、ある音程行ベクトルをもつピッチクラス集合は何かというように音程が先に決定される場合がある。これらのケースは、音高よりも音程を優先した見方だと考えることができる。

このように、音高と音程のどちらを優先させるかによって、同じ対象を異なる角度から見ることができ、それによってより豊かな音楽の解釈がもたらされるだろう。図 4.1 に音高と音程の両観点からの音楽概念をまとめる。

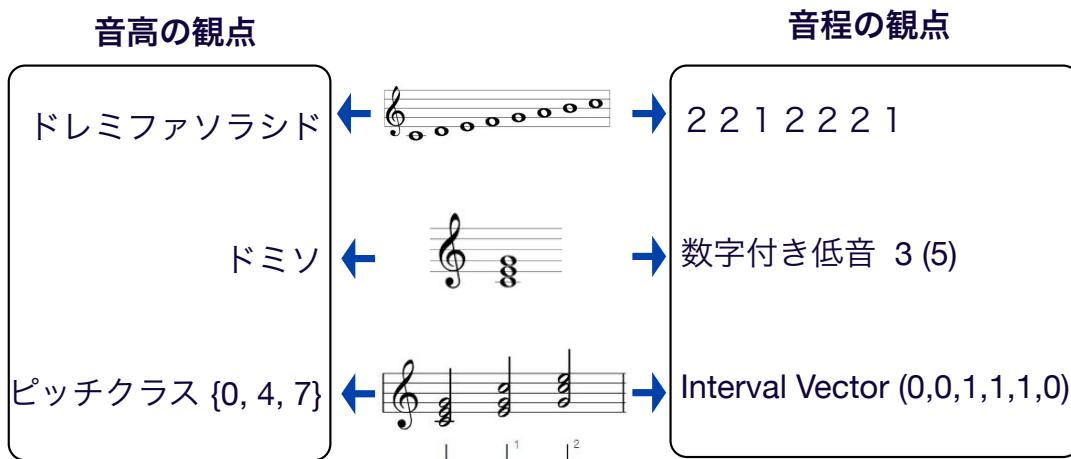


図 4.1. 音高および音程の観点から見た音楽概念。ピッチクラスの集合およびその音程行ベクトルは、和音の転回形の違いによらず同一である。

だが、両者の見方のうち、多くの人にとってなじみ深いのは音高を重視する見方であるように思われる。これはなぜだろうか。私はその理由が、音程からの見方が、ある意味で把握しにくかったり、曖昧性を持っているからではないかと考える。例えば、図 4.1 右上の音階の隣接音程による表示は、ドからラへの跳躍などといった隣接しない場合の音程を陽に記述していたために直接的に把握しにくい。また音程行ベクトルの概念は、同じベクトルをもつ異なるピッチクラス集合が存在する場合があるという曖昧性がある。それゆえに、音高からの見方よりも認知的負荷が高く、音楽理論としてマイナーになりがちなのではないかと思われる。しかし、音程スケールの場合には、形式的には音階と同様に \mathbb{Z}_n のある部分集合として定義されるため、音高の観点からの概念と比較しても明晰さにおいて劣らず、使い勝手の良い概念だと考えられる。

では、音程スケールが明晰な概念であるのは良いとしても、実際の音楽の分析や作曲におい

て有用な概念たりえるだろうか？この点は、音程スケールを適用できる既存曲が存在するかどうか、あるいは今後作曲される新しい楽曲において有効に用いられるかどうかにかかっている。

以下では、音程スケールの概念の可能性を検証するため、次のような流れで考察を行う：4.4節では音程スケールに基づいていると解釈できるような楽曲が存在することを実例を挙げて示す。その後4.5節では、音程スケールの一般的な構造を探求し、音程スケールが群という代数的構造と密接な関係があることを示す。そして4.6節では、音程スケールをどう選択すれば良いかという指針となる重要な2つの定理を示す。4.7節では、12音列を含む一般の音列に対して音列操作のなす群を定義し、音列操作が音程スケールに与える影響関係について議論する。そして最後に、それまで考察した理論に基づいて音程スケールを選択する手順の一例を示し、選択された音程スケールを用いて著者が作曲した楽曲について述べる。

4.4 音程スケールの概念を適用できる楽曲例

本節では音程のスケールに基づいていると解釈できるような2つの楽曲例を取り上げ、音程スケールの作曲への応用における有用性を示す。

4.4.1 Ligeti 「エチュード第2番（開放弦）」

Ligeti の作品『ピアノのためのエチュード第2番「開放弦 (Cordes à vide)」』[34] は「開放弦」というタイトルからもわかるように、完全5度音程の上下行 ($\bar{7}$ と $\bar{5}$ の音程) を中心に楽曲が作られている。図4.2にこの曲の冒頭を示す。

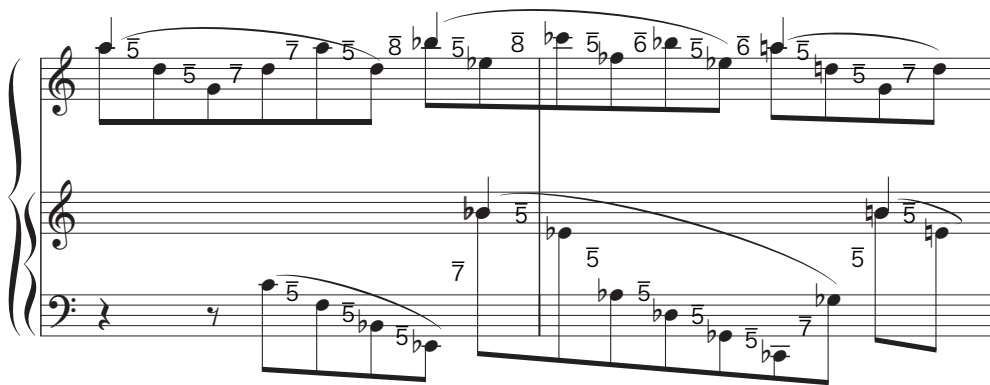


図4.2. Ligeti 「エチュード第二番」冒頭。数字は隣接音の間の半音を単位とした音程を表す。

第2小節目の終わりまでの8分音符の旋律的音程を全て集めると $\{\bar{5}, \bar{6}, \bar{7}, \bar{8}\}$ という音程スケールが構成される。これは要素数が少なく、音程が連続しているという特徴をもつ。音程 $\bar{6}$ および $\bar{8}$ は $\bar{5}$ と $\bar{7}$ にそれぞれ隣接することで、完全5度の動きを緩やかにトランスポートすることを可能にし、同じピッチクラスへの膠着を防ぐ役割を担っていると考えられる。他方、

ピッチに関しては、右手パートの二つ目のスラーの終わりの時点までには全てのピッチクラスが早くも出現するため、クロマティックスケール \mathbb{Z}_{12} に基づくと解釈できる。音程スケールに基づくとき、音高のスケールの観点からするとクロマティックスケールになるのは偶然ではない。なぜなら例えば $\bar{7}$ のみの音程だけでも、五度圏をたどることで全てのピッチクラスが出現しうるからである。逆に言えば、クロマティックスケールに基づく無調音楽であっても、実は要素数の小さい音程スケールが使用され、偏りのある響きをもたらされているケースが想定されうる。この曲も「開放弦」というサブタイトルが示すように、音程の限定による響きの偏りは多かれ少なかれ意図的だと考えられ、ピッチにおいて無調であるにもかかわらずある種の輝きのある調性感のようなものを有している。ここでは音程スケールが効果を発揮していると考えてよいだろう。

4.4.2 Berg 「ヴァイオリン協奏曲」の12音音列

柴田南雄は文献 [35] において、Schönberg が 12 音技法や無調という言葉を使っておらず、「相互の間の関係のみに依存している 12 個の音による作曲の方法」と称していたことに注意を向けている。そして、12 音音楽といえども調性感をも積極的に包含する可能性があり、12 音音列の機械的適用を超えた未知の秩序を見いだしうるというスタンスで 12 音音楽を論じている。また小鍛冶邦隆は文献 [36] において、調性音楽と無調音楽の接点における創作の方法を探るため、Berg の「ヴォツェック」の分析を行っている。ここで手がかりになっているのは「調的要素」としての音程である。このように、音程の分析は無調音楽の中の調性感を理解する一つの手がかりになると考えられる。

ここでは、12 音音楽における調性感という一見矛盾する事象を良く体現する楽曲として有名な Berg の「ヴァイオリン協奏曲」[37] の 12 音音列をとりあげる (図 4.3)。

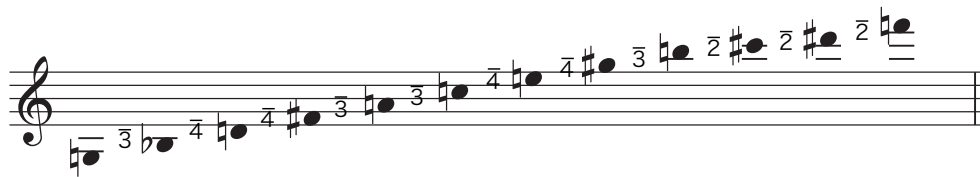


図 4.3. Berg 「ヴァイオリン協奏曲」の 12 音音列 [38].

この音列の構成法は容易に見てとれる。始めは長三度と短三度の音程が規則的に現れ、最後の三つの音程が全音となっている。これらの音程を集めた集合は $\{2, 3, 4\}$ という音程スケールをなす。長短の三度音程や、全音の音程は、調性音楽の特徴を表す重要な音程であり、それがこの曲の調性感につながっていると考えられる。

音程の要素数の少ないこの音列とは対照的に、 $\bar{0}$ を除く 11 種類全ての音程が全て出現する音列は総音程音列と呼ばれる [39]。これは音程スケール $\{\bar{1}, \bar{2}, \bar{3}, \bar{4}, \bar{5}, \bar{6}, \bar{7}, \bar{8}, \bar{9}, \bar{10}, \bar{11}\}$ に基づく音列としてとらえ直すことができる。総音程音列においては、ピッチだけでなく音程につい

でもできるだけ多くの種類を偏りなく用いることで、調性感が逆に排除されると考えられる。総音程音列のような概念が存在することは、ピッチクラス集合の固定された12音音列中から音程の選択の自由度があることを意味している。先に挙げた Ligeti と Berg の例は、この自由度を活かして音程を組織することによって調性感のある響きを作り出していると解釈でき、音程スケールは無調性と調性感を媒介する役割を担いうると思うことができる。また、Ligeti と Berg の楽曲例では、前者は比較的明るく、後者は比較的暗い響きを持っており、音程スケールの違いが異なる響きを生み出していると解釈できる。

以上のように、音程スケールは実際の音楽で有効に働きうるということがわかる。

4.5 \mathbb{Z}_n の代数的構造

前節では実際の楽曲の背後に音程スケールを見だし、その働きを考察した。次に出てくる問題は、新しい曲を作る際にどのように音程スケールを適切に選択するかということである。例えば Ligeti 作品の音程スケール $\{\bar{5}, \bar{6}, \bar{7}, \bar{8}\}$ や Berg 作品の音程スケール $\{\bar{2}, \bar{3}, \bar{4}\}$ は要素数が少なく、連続しているという特徴を共有する。音程スケールの選び方は多数ある中で、これらの選択にはどのような意味があるのだろうか？それを理解するための準備として、本節では音程のシステムとしての \mathbb{Z}_n の中での音程スケールのありようを数学的に考察する。

4.5.1 音程の加法的性質

1 オクターブ n 音の調律システムにおけるピッチクラス集合およびピッチ間の音程のクラスの集合は共に $\mathbb{Z}_n = \{\bar{0}, \bar{1}, \dots, \overline{n-1}\}$ になる。数学的には \mathbb{Z}_n は単なる要素の集合であるだけでなく、任意の2つの要素の足し算が定義された群という代数的構造を持つ。しかし、 \mathbb{Z}_n をピッチクラスの集合としてみたときには、足し算の演算に音楽的な意味を見いだすことは困難である。つまりピッチクラス A とピッチクラス B を足した結果が別のピッチクラスを表すというような足し算（例えば $n = 12$ のとき、 $\bar{2} + \bar{10} = \bar{0}$ ）を考えても、そこに何の意味があるのかが解釈しにくい。他方、 \mathbb{Z}_n を音程のクラスの集合としてみると、音程 A と音程 B を足すと音程 C になる（例えば 3 半音 + 4 半音 = 7 半音）というように、足し算が自然な意味をもつ。これは時刻と時間の違いにも類似する事象である。つまり「2 時と 10 時をたす」ということには意味が無いが、「2 時間 + 10 時間 = 12 時間 = 0 時間（時計を一周して元通り）」ということには意味があるのと同様の事態である。したがって群論的には、ピッチクラスの集合ではなく、音程のクラスの集合としての \mathbb{Z}_n が興味の対象である。

4.5.2 群の生成系としての音程スケール

ここで群論の基本概念的定義を確認する [40]。

定義 2 (群 (group)). 集合 G の任意の元 a, b に対し、ある二項演算 ab が定義されており、 $ab \in G$ (演算について閉じている) とする。このとき G が群であるとは、この演算が次の三

条件を満たすときにいう：

1. G の演算は結合法則を満たす. すなわち, 任意の元 $a, b, c \in G$ に対して $(ab)c = a(bc)$ が成り立つ.
2. G に単位元が存在する. すなわち, ある $e \in G$ が存在し, 任意の $a \in G$ に対し, $ae = ea = a$ となる.
3. G の任意の元に対して逆元が存在する. すなわち, 任意の $a \in G$ に対し, $ab = ba = e$ となる $b \in G$ が存在する.

\mathbb{Z}_n は群の典型的な例の一つである. $\bar{a}, \bar{b} \in \mathbb{Z}_n (0 \leq a, b \leq n-1)$ とすると, $a+b$ を n で割った余り c によって, $\bar{a} + \bar{b} = \bar{c}$ という \mathbb{Z}_n の中で足し算 “+” が自然に定義できる. このとき単位元は $\bar{0}$, \bar{a} の逆元は $\overline{n-a}$ (ただし $\bar{n} = \bar{0}$) である. この逆元は $-\bar{a}$ と表記する.

群 G の部分集合 H がそれ自体群になっているとき H を部分群という. 群 G の元の個数を G の位数という. また, 集合 G が群の定義において第 2, 第 3 の条件を満たすとは限らないとき G は半群であるという [41]. つまり, 半群は単位元と逆元を必ずしも持つ必要が無い. 半群 G の部分集合 H がそれ自体半群になっているとき H を部分半群という.

定義 3 (群の生成系 (group generators)). 群 G の部分集合 S を含む最小の部分群を S の生成する群といい $\langle S \rangle$ で表す. S は $\langle S \rangle$ の生成系という.

+ の演算を用いると, S の生成する群 $\langle S \rangle$ は次式のように構成的に表示できる.

$$\langle S \rangle = \{\epsilon_1 s_1 + \epsilon_2 s_2 + \cdots + \epsilon_r s_r \mid s_j \in S, \epsilon_j = \pm 1, r \in \mathbb{N}\}. \quad (4.1)$$

$\langle S \rangle$ は必ず存在し, $\langle S \rangle$ は S を含み, かつ S を含む任意の群に含まなければならない集合である. また, この集合が群になることはすぐに定義から確認できる. したがってこれは S を含む最小の部分群であることがわかる. この式は, $\langle S \rangle$ の全ての要素は, S の要素とその逆元の組み合わせで表現できることを意味する. 他方, 半群の生成系も同様に定義できる:

定義 4 (半群の生成系 (Generators of semigroup)). 半群 G の部分集合 S を含む最小の部分半群を S の生成する半群といい $\ll S \gg$ で表す. S は $\ll S \gg$ の生成系という.

+ の演算を用いると, S の生成する半群 $\ll S \gg$ は次のように構成的に表示できる:

$$\ll S \gg = \{s_1 + s_2 + \cdots + s_r \mid s_j \in S, r \in \mathbb{N}\}. \quad (4.2)$$

これは $\ll S \gg$ の全ての要素は S の要素の組み合わせで表現できることを意味する. 群の生成と半群の生成における違いは S の逆元を用いるかどうかである.

音程スケールを用いて旋律 (音列) を作ることは, 半群の生成において, 式 (4.2) 右辺の各要素を, 音程スケールを S とした音程の足し算によって逆元を用いずに次々と作り出していくことに対応する. 図 4.4 は同じ音程スケールから複数の旋律が作られるさまを表す. ここでは音程スケール $\{\bar{2}, \bar{6}\}$ (これは式 (4.2) の S に対応する) の要素が自由に選択され, 逐次的に足されて旋律が作られていくさまを表している.

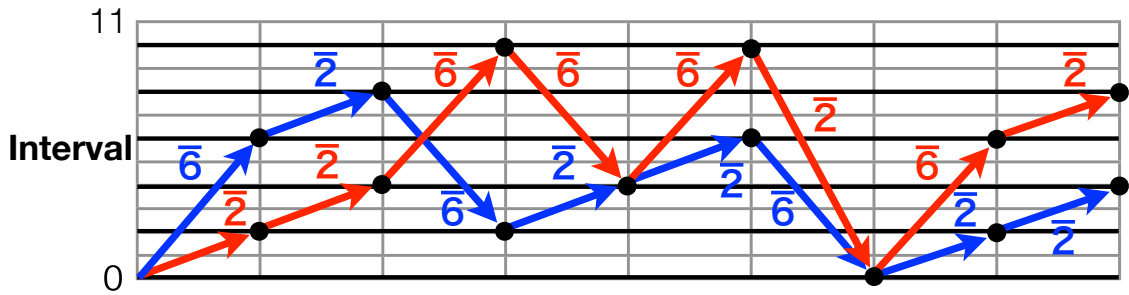


図 4.4. 同じ音程スケール $\{\bar{2}, \bar{6}\}$ から作られる二つの旋律（赤と青のライン）と生成される半群 $\langle\langle \bar{2}, \bar{6} \rangle\rangle = \{0, \bar{2}, \bar{4}, \bar{6}, \bar{8}, \bar{10}\}$ （黒丸の全体）.

式 (4.2) は、図 4.4 のようにして、音程スケールから全ての可能な旋律を作るプロセスにおいて登場しうるすべてのピッチクラス ($\bar{0}$ を基準とした音程) を集めた集合が半群を形成するさまを表すものと解釈できる. 先の Ligeti と Berg の譜例はそれぞれ音程スケール $\{\bar{5}, \bar{6}, \bar{7}, \bar{8}\}$ と $\{\bar{2}, \bar{3}, \bar{4}\}$ から半群としての \mathbb{Z}_{12} 全体が逐次的に生成されるプロセスを表しているのととらえ直す事ができる.

ところで、式 (4.1) と (4.2) を比較すると、一般に $\langle S \rangle$ は $\langle\langle S \rangle\rangle$ より大きいか一致するかのいずれかであることがわかる. つまり、 S の要素の逆元を用いて生成された群は、逆元を用いないで生成された半群を含む. しかし、 S が \mathbb{Z}_n の部分集合としての音程スケール (これを I_n と表す) である場合には、 I_n によって生成される群と半群は一致する (以後 I_n を \mathbb{Z}_n のある部分集合としての音程スケールを表すものとする). すなわち次の命題が成り立つ:

補題 1. 半群 $\langle\langle I_n \rangle\rangle$ は群 $\langle I_n \rangle$ に一致する.

証明. 既に述べたように \mathbb{Z}_n は巡回群である. したがって、巡回群の定義より、位数 $n(n \in \mathbb{N})$ の巡回群 G の任意の元 s は $ns = e$ を満たす. また $(n-1)s + s = ns = e$ より $(n-1)s$ は s の逆元 $-s$ になる. $(n-1)s \in \langle\langle I_n \rangle\rangle$ なので、 $\langle\langle I_n \rangle\rangle$ には S の任意の元の逆元が含まれる. したがって $\langle I_n \rangle$ の任意の元は $\langle\langle I_n \rangle\rangle$ の元として書き表すことができ、 $\langle I_n \rangle \subseteq \langle\langle I_n \rangle\rangle$ となる. また $\langle I_n \rangle \supseteq \langle\langle I_n \rangle\rangle$ は構成的な表示による定義より明らかである. したがって $\langle I_n \rangle = \langle\langle I_n \rangle\rangle$ となる. □

この補題より、半群 $\langle\langle I_n \rangle\rangle$ の群としての生成系は音程スケール I_n だと考えて良い. また半群 $\langle\langle I_n \rangle\rangle$ が群になるということは、音楽的には、逆元によって音程移動の後戻りができることが保証されたことを意味する.

4.5.3 巡回群と五度圏

\mathbb{Z}_n は n 個の要素をもち、それらのあらゆる元は $\bar{1}, \bar{1} + \bar{1} = \bar{2}, \bar{1} + \bar{1} + \bar{1} = \bar{3}, \dots, \bar{1} + \bar{1} + \dots + \bar{1} = \overline{n-1}$ というように、 $\bar{1}$ という最小単位の音程上昇を表す一つの元を一つづつ足していくことで尽くされる. そして、 n 個足したとき、最初の $\bar{0}$ という地点に戻り、循環

する. このように, 群 G が一つの元 $g \in G$ の n 個までの和の循環によって尽くされる群は位数 n の巡回群と呼ばれる. そしてこの g は巡回群の生成元と呼ばれる.

一般に, 生成元の取り方は $\bar{1}$ だけとは限らない. 例えば 12 平均律における五度圏は, 完全 5 度音程を表す元 $\bar{7}$ によって $\{\bar{0}, \bar{7}, \bar{2}, \bar{9}, \bar{4}, \bar{11}, \bar{6}, \bar{1}, \bar{8}, \bar{3}, \bar{10}, \bar{5}, \bar{0}\}$ というように \mathbb{Z}_{12} が巡回群として逐次的に生成されていく過程と解釈できる. しかし, 全音を表す音程 $\bar{2}$ を生成元にとろうとすると, $\{\bar{0}, \bar{2}, \bar{4}, \bar{6}, \bar{8}, \bar{10}\}$ と, 偶数しか出現しないため \mathbb{Z}_{12} 全体をカバーできず, 生成できない. しかし, 部分集合としての $\{\bar{0}, \bar{2}, \bar{4}, \bar{6}, \bar{8}, \bar{10}\}$ はそれ自体, 群となり (部分群), $\bar{2}$ はその生成元となっている.

生成元のとりにかたしだいで巡回群の要素数は, 例えば図 4.5 (a) (12 個) と (b) (2 個) のように変化する. また, n が変化すれば同じ生成元から生成される巡回群の要素数も変化する. 例えば図 4.5 の (a) と (c) のように, $n = 12$ のときに $\bar{7}$ は \mathbb{Z}_{12} 全体を生成するが, $n = 21$ のときに同じ $\bar{7}$ は \mathbb{Z}_{21} 全体を生成しない.

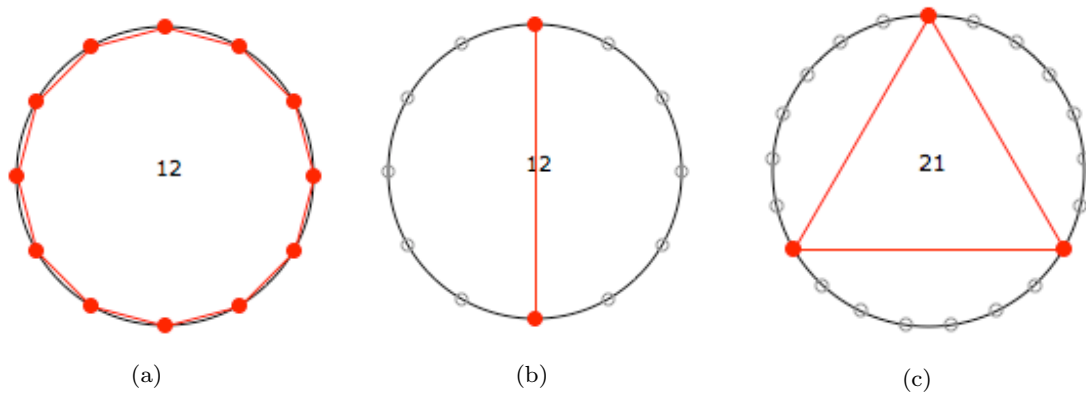


図 4.5. n および生成元を変えた時の巡回群の変化. (a): $n = 12$, 生成元が $\bar{1}, \bar{5}, \bar{7}, \bar{11}$ のいずれかるとき. (b): $n = 12$, 生成元が $\bar{6}$ のとき. (c): $n = 21$, 生成元が $\bar{7}$ または $\bar{14}$ のとき.

4.5.4 五度圏の二段階の拡張

興味深いことに, Berg は Schönberg への手紙 [42] の中で, 五度圏のように特定の音程を積み重ねて反復することで元の音に戻ってくるサイクルの一覧を描いており, Berg のオペラ作品「ヴォツェック」でしばしば素材として用いられてもいる [36]. これらのサイクルのことを作曲家 George Perle はインターバル・サイクル (Interval cycles) と呼んでいる. これらは一つの生成元から生成される巡回群の一覧になっている. このように巡回群 (インターバル・サイクル) は五度圏の拡張とみなせる.

音程スケールの観点からは, インターバル・サイクルはさらに拡張できる. インターバル・サイクルの場合は生成系はただ 1 つの元からなり, ただ 1 つの旋律しか生み出す事ができないが, 複数の要素を持つ生成系からは, 図 4.4 のように複数 (多数) の旋律を生み出す事ができ

る。このように、音程スケールからの群の生成は、インターバル・サイクルの生成系の要素数を複数化したさらなる拡張と見なせる。そして、この拡張によって、音階からピッチの自由に選択できるのと同じように音程スケールを使用することが可能になる。つまり、どの音程をどの順に積み重ねていくかを自由に選択できるようになり、その結果、様々な旋律を作り出すことができるようになる。

4.6 音程スケールの選択方法

音程スケールは \mathbb{Z}_n の部分群を生成する。では、それはどのような部分群だろうか？²幸いにも、巡回群の部分群は巡回群であることが知られている [40, 43]。したがって \mathbb{Z}_n が巡回群であることから、 \mathbb{Z}_n の部分群は巡回群であることがわかる。次の問題は、音程スケールの取り方と、そこから生成される巡回群の間にはどのような関係があるかを解明することである。音程スケールの取り方によって、図 4.4 のような位数の小さな部分群になることもあれば、Ligeti や Berg の例のように、 n 種類の音程全てを含む \mathbb{Z}_n 全体が生成されることもあるだろう。また、本論文では音程スケールを無調音楽において響きをコントロールするシステムとして位置づけているため、どのような時に \mathbb{Z}_n 全体が生成されるのか（つまり無調になるのか）ということが特に問題である。これが解明されれば、音程スケールを用いて作曲する際に音程スケールの選択の指針となるだろう。以下ではこの問題を扱う。

まず、無調性の基準として次の用語を定義しておく：

定義 5 (クロマティック (Chromatic)). I_n を調律システム \mathbb{Z}_n におけるある音程スケールとする。 $\langle I_n \rangle = \mathbb{Z}_n$ のとき生成される群 $\langle I_n \rangle$ はクロマティックであるという。

次の補題は有名なユークリッドの補題の一般化であり、後の定理の証明の鍵となる。

補題 2 (ユークリッドの補題の一般化). $i > 0$ に対して y_i は整数とし、 $\{y_1, y_2, \dots, y_i\}$ のすべての元の最大公約数 $\gcd(y_1, y_2, \dots, y_i)$ を d とする。このとき $a_1 y_1 + a_2 y_2 + \dots + a_i y_i = d$ となるような整数の組 a_1, a_2, \dots, a_i が存在する。ただし、 $\gcd(y_1) = |y_1|$ とし、 $\gcd(0, 0, \dots, 0) = 0$ と約束する。

証明. この命題は、ユークリッドの補題と呼ばれる有名な命題を一般化したものである。数学的帰納法で命題を証明する。 $i = 1$ のときは $\gcd(y_1) = |y_1|$ より $a_1 = 1$ または $a_1 = -1$ とすれば命題が成り立つことがわかる。 $i = 2$ のとき、命題は通常のユークリッドの補題に一致するため命題は成り立つ。いま、 $i = k (k \leq 1)$ のとき命題が成り立つと仮定する。 $i = k + 1$ のときの命題の前提が成り立っているとすると $\gcd(y_1, y_2, \dots, y_k, y_{k+1}) = d$ 。これにより、

² 音程のなす群についてはこれまでに多くの研究がなされてきた。例えば、Xenakis は音程のなす群に言及している [44]。また Lewin の GIS (generalized interval system) はピッチクラスだけではなく、一般の音楽的要素の空間のインターバルの群 IVLS を考えた [45, 46]。Morris は音列の操作のなす群について詳しく研究した [47]。しかし、これらの研究においては、群が複数の要素をからなる生成系からどのように生成されるのかについてあまり探求されていない。

$\gcd(y_1, y_2, \dots, y_k)$ を D とおくと, $\gcd(y_1, y_2, \dots, y_k, y_{k+1}) = \gcd(D, y_{k+1}) = d$ である. したがって $k=2$ のときの命題により, $AD + By_{k+1} = d$ となる整数の組 A, B が存在する. また, $i=k$ のとき命題が成り立つので, $(a_1y_1 + a_2y_2 + \dots + a_ky_k) = D$ となる a_1, a_2, \dots, a_k がとれる. 左辺を $AD + By_{k+1} = d$ の D に代入すると, $A(a_1y_1 + a_2y_2 + \dots + a_ky_k) + By_{k+1} = d$ となる. ここでの y_i の係数をみれば, $i=k+1$ のとき命題が成り立っていることがわかる. 以上, 数学的帰納法により, 任意の $i \geq 1$ に対して命題が証明された. \square

補題を用いて二つの重要な定理を示す. また, 第二の定理から導かれる系として, $\langle I_n \rangle$ がクロマティックになる条件に関わる実用上重要な帰結を示す.

定理 1 (最大公約数による巡回群). $0 \leq x_k \leq n-1 (1 \leq k \leq i, i > 0, x_k \in \mathbb{Z})$, $I_n = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_i\}$ のとき, $\langle I_n \rangle$ は $\{x_1, x_2, \dots, x_i\}$ の最大公約数 d のクラス \bar{d} によって生成される巡回群 $\langle \bar{d} \rangle$ に一致する.

証明. $\{x_1, x_2, \dots, x_i\}$ の全ての元は d の倍数だから $\langle IS_n \rangle \subseteq \langle \bar{d} \rangle$ であることはすぐわかる. 次に $\langle I_n \rangle \supseteq \langle \bar{d} \rangle$ であることを示す. そのためには $\bar{d} \in \langle I_n \rangle$ であることを示せば良いが, I_n の元 x_1, x_2, \dots, x_i の最大公約数を d とすると, 補題 2 より $a_1x_1 + a_2x_2 + \dots + a_ix_i = d$ となる整数の組 a_1, a_2, \dots, a_i がとれる. よって $a_1\bar{x}_1 + a_2\bar{x}_2 + \dots + a_i\bar{x}_i = \bar{d}$ となり, 左辺は $\langle I_n \rangle$ の構成的な表示式で表現できるから $\bar{d} \in \langle IS_n \rangle$ が成り立つ. \square

この定理は, $\langle I_n \rangle$ が \bar{d} によって定まり, 分類できることを示す. 例えば \mathbb{Z}_{14} において, $\{\bar{0}, \bar{8}, \bar{10}\}$ および $\{\bar{6}, \bar{8}\}$ という両音程スケールは最大公約数 2 を持つため, 同じ $\langle \bar{2} \rangle = \{\bar{0}, \bar{2}, \bar{4}, \bar{6}, \bar{8}, \bar{10}, \bar{12}\}$ という部分群の生成系であることがわかる.

定理 2 (位数公式). $\langle I_n \rangle (= \langle \bar{d} \rangle)$ の位数は $\frac{n}{\gcd(d, n)}$ である.

証明. xd が n の倍数になる最小の $x (\geq 1)$ が \bar{d} の位数である. そこで, $xd = kn$ とおく ($k \geq 0$). $d \neq 0$ のとき, $d = d_1 \gcd(d, n)$, $n = n_1 \gcd(d, n)$ とおくと, $x = \frac{kn_1}{d_1}$ である (ただし d_1 と n_1 は互いに素である). x が整数であり, d_1 と n_1 は互いに素であることから, k は d_1 の倍数でなければならないが, x の最小性より $k = d_1$ でなければならない. したがって, $x = n_1 = \frac{n}{\gcd(d, n)}$. $d = 0$ のときは $k = 0$ であり, $xd = kn$ を満たす最小の $x (\geq 1)$ は 1 である. また $\gcd(0, n) = n$ だから公式が成り立つ. \square

この定理から, 部分群がクロマティックになる条件に関する次の二つの系が導かれる.

系 1 (クロマティックになる条件). $\langle I_n \rangle$ がクロマティックになる必要十分条件は d が n と互いに素であることである.

証明. 定理 2 より, I_n がクロマティック $\iff \gcd(d, n) = 1 \iff d$ が n と互いに素. \square

この系より、特に n が素数で $d > 0$ のときは、 d と n が互いに素なので $\langle I_n \rangle$ が常にクロマティックになることがわかる。

系 2 (連続する 2 元を含む場合). I_n が連続する二つの元 $\bar{x}, \overline{x+1}$ ($x = 0, 1, \dots, n-1$) を含むとき、 $\langle I_n \rangle$ はクロマティックになる。

証明. x と $x+1$ は互いに素だから $d = 1$. よって定理 2 から $\langle I_n \rangle$ の位数は n となり、 $\langle I_n \rangle$ は \mathbb{Z}_n に一致する。□

この系は、先に挙げた Ligeti と Berg の例のように、連続する音程を含む I_n が常にクロマティックな群を生成することを示している。

これら二つの系は、 I_n が全てのピッチクラスを使用する無調音楽を生成できるかどうかを判定するための目安として用いることができる。

系 1 の帰結である n が素数で $d > 0$ のとき $\langle I_n \rangle$ が常にクロマティックになることは極めて示唆的である。なぜなら通常の調律システムの音数である 12 という数は素数ではないが、両隣が 11 と 13 という素数であり、あたかも素数になることを避けているかのようだからである。さらに、素数とは対照的に 12 という数は d の取り方によって生成されうる部分群のバリエーションが多いという特徴をもつ (定理 2 から位数が 1, 2, 3, 4, 6, 12 の 6 通りであることがわかる)。これは 12 が約数を多く持つことに起因する。実際 24 より小さい自然数で約数の個数が最も多いのは 12 であることを考えると、調律システムの音数に 12 が選ばれるのは偶然ではないのではないかという仮説を立てることができる。この考えに基づくならば、一オクターブ n 音の調律システム \mathbb{Z}_n における n の値を決定するための 1 つの指針が得られる。すなわち、 n の約数の個数がある n の範囲内で最大にする n を選択することで、良い調律システムが選択できるという考え方をとることができる。

また、系 2 からは次のことがいえる。音程スケールの要素数が少ないことは、響きの特徴や調性感を明確化するのに役立つ一方で、最大公約数が 1 になりにくいために生成する群がクロマティックになりにくいことを意味する (定理 1 参照) が、系 2 における連続性を導入することで無調性を保証することができる。したがって、連続な音程スケールを用いた Ligeti と Berg の選択は、無調性と少ない要素数を両立させる効率的な戦略だったのだと解釈することができる。

最後に参考までに、得られた定理を用いて、12 平均律の場合の音程スケールの生成する群を d で分類してみる。系 1 より、 d が n と共通因数を持たないとき、すなわち $d = 1, 5, 7, 11$ のとき音程スケールはクロマティックである。残る $d = 2, 3, 4, 6, 8, 9, 10, 12$ のときはクロマティックにならない。それぞれの場合の位数は定理 2 の公式より 6, 4, 3, 2, 3, 4, 6, 1 である。また定理 1 より、音程スケールの生成する群 I_{12} は次のように分類される：

- $d = 1, 5, 7, 11$: クロマティックな音程スケール Z_{12} .
- $d = 2, 10$: 全音音程スケール $\{\bar{0}, \bar{2}, \bar{4}, \bar{6}, \bar{8}, \bar{10}\}$.
- $d = 3, 9$: 短三度音程スケール $\{\bar{0}, \bar{3}, \bar{6}, \bar{9}\}$.

- $d = 4, 8$: 三度音程スケール $\{\bar{0}, \bar{4}, \bar{8}\}$.
- $d = 6$: 三全音音程スケール $\{\bar{0}, \bar{6}\}$.
- $d = 12$: 完全一度音程スケール $\{\bar{0}\}$.

4.7 音列操作と音程スケール

12 音音列に対する, 順行 (e), 逆行 (i), 反行 (r), 逆行の反行 ($ir = ri$) という操作は, クラインの四元群という位数 4 の群と同型であることは良く知られている. この群は, 巡回群ではない群の中で最も位数の小さなものである. 本節では, これらの操作を n 音の調律システムにおける一般の音列に対する操作として改めて定義し, 音列操作が, 音程スケールに及ぼす影響を明らかにする.

定義 6 (音列の音程スケール). n 音の調律システム \mathbb{Z}_n における m 音の音列を, あらゆる (n, m) ($n \in \mathbb{N}, m > 1$) の組み合わせについて集めた集合を X , あらゆる $n \in \mathbb{N}$ に対する \mathbb{Z}_n の音程スケールをすべて集めた集合を Y とする (X, Y において n が異なるが実質的に同じ元の場合は同一視する). 写像 $IS : X \rightarrow Y$ を, 音列 $x \in X$ において x 内の各音から次の音への音程を全て集めた集合 (要素の重複を除いた集合) に対応付ける写像とする. このとき, $IS(x)$ を音列 x の音程スケールと呼ぶ.

定義 7 (順行, 逆行, 反行, K). $Map(X, X)$ を集合 X からそれ自身への写像全体の集合とする. このとき, e, i, r を $Map(X, X)$ の元として, e を恒等写像, i を音列 $x \in X$ をその逆行 (ただし音列の最初のピッチと最後のピッチの中間点で折り返す) に対応させる写像, r を音列 x をその反行に対応させる写像と定義する. $m_1, m_2 \in Map(X, X)$ の写像の合成を $m_1 m_2$ とすると, $ii = e, rr = e$ (反行の反行および逆行の逆行は順行) および, $ir = ri$ (逆行の反行は反行の逆行に等しい) が成り立つ. これは反行, 逆行の性質上明らかである. そして e, i, r, ir は写像の合成の演算に関してクラインの四元群になる (e は単位元, i, r はそれ自身が逆元であり, $(irir = irri = iei = ii = e$ より ir の逆元はそれ自身). この群を K と表記する. また K の元 k によって音列 x を写像した先を $k \cdot x$ として, 演算記号 \cdot を用いて表す.

定義 8 (音程スケールの反転 inv, B). $Map(Y, Y)$ の元としての inv を, \mathbb{Z}_n に所属する任意の音程スケール $y \in Y$ に対して, y のの各要素音程 $\bar{y}_i \in \mathbb{Z}_n$ を $\overline{n - y_i}$ に置き換えた音程スケールに対応付ける写像とする. また, E を $Map(Y, Y)$ の恒等写像とする. このとき, 写像 inv を, 音程スケールの反転とよぶ. inv は二回繰り返すと元に戻る. したがって, $\{E, inv\}$ は写像の合成の演算に関して位数 2 の巡回群となる (E は単位元で inv の逆元はそれ自身). この群を B と表記する. また B の元 b によって音程スケール y を写像した先を $b \circ y$ として, 演算記号 \circ を用いて表す.

定義 9 (K から B への準同型写像 f). 写像 $f : K \rightarrow B$ を以下のように定義する: $f(e) = E, f(i) = inv, f(r) = inv, f(ir) = E$. このとき f は準同型写像である. すなわち, 任意の

$k_1, k_2 \in K$ に対し, $f(k_1 k_2) = f(k_1) f(k_2)$ が成り立つ (このことは B における E と inv の定義からすぐに確認できる).

定理 3 (音列操作と音程スケールの操作の関係). 任意の音列 $x \in X$ に対して次の四つが成り立つ:

1. 順行の音程スケールはもとの音程スケールに一致する; $IS(e \cdot x) = E \circ (IS(x))$.
2. 反行の音程スケールはもとの音程スケールの反転に一致する; $IS(i \cdot x) = inv \circ (IS(x))$.
3. 逆行の音程スケールはもとの音程スケールの反転に一致する; $IS(r \cdot x) = inv \circ (IS(x))$.
4. 逆行の反行の音程スケールはもとの音程スケールに一致する; $IS((ir) \cdot x) = E \circ (IS(x))$.

すなわち, まとめると, 任意の $k \in K, x \in X$ に対して $IS(k \cdot x) = f(k) \circ (IS(x))$ であり, 次の可換図式が成り立つ:

$$\begin{array}{ccc}
 K \times X & \xrightarrow{f \times IS} & B \times Y \\
 \downarrow \cdot & & \downarrow \circ \\
 X & \xrightarrow{IS} & Y
 \end{array}$$

証明. まず準備として, 音列 $x \in X$ に IS を施す操作を次の三つの手順, α, β, γ の合成として表す: 手順 1. $x = [x_1, x_2, \dots, x_j]$ から隣接する要素の差のリスト $\alpha(x) = [x_2 - x_1, x_3 - x_2, \dots, x_j - x_{j-1}]$ を作る. 手順 2. $\alpha(x)$ の各要素を \mathbb{Z}_n の元に変換する. これを $z = \beta(\alpha(x)) = [\overline{x_2 - x_1}, \overline{x_3 - x_2}, \dots, \overline{x_j - x_{j-1}}]$ と表記する. 手順 3. リスト $\beta(\alpha(x))$ から重複する要素を省いた集合を作る. これを $\gamma(z)$ とする. すなわち, $IS(x) = \gamma(\beta(\alpha(x))) = \gamma(z)$ である.

1. の証明: $IS(e \cdot x) = IS(x)$, $E \circ (IS(x)) = IS(x)$ であるから 1. が成り立つ.

2. の証明: $\alpha(i \cdot x) = [-(x_2 - x_1), -(x_3 - x_2), \dots, -(x_j - x_{j-1})]$ だから, $\beta(\alpha(i \cdot x))$ は z の各要素をその逆元に置き換えたリストである. これを w と置くと, 2. の左辺 = $\gamma(w)$ である. 2. の右辺は $inv \circ (IS(x)) = inv \circ (\gamma(\beta(\alpha(x)))) = inv \circ (\gamma(z))$ であるが, inv は定義上, 集合の各要素の逆元を集める操作であり, 要素に重複があるリストに対しては最後に重複を除いた集合をとるといのように定義を拡張すると, $inv \circ (\gamma(z))$ は $inv \circ (z) = \gamma(w)$ に一致する, したがって 2. の左辺 = 右辺が成り立つ.

3. の証明: $\alpha(r \cdot x) = [x_{j-1} - x_j, x_{j-2} - x_{j-1}, \dots, x_1 - x_2] = [-(x_j - x_{j-1}), -(x_{j-1} -$

$x_{j-2}), \dots, -(x_2 - x_1)]$ だから, 2. の証明より, これは $\alpha(i \cdot x)$ の各要素を逆順に読んだリストである. これを $\gamma(\beta())$ で操作して集合をとるとき, 要素の順番は無関係になる. したがって 3. の左辺 $= \gamma(\beta(\alpha(r \cdot x))) = \gamma(\beta(\alpha(i \cdot x))) = IS(i \cdot x) = inv \circ (IS(x))$.

4. の証明: $\alpha((ir) \cdot x) = [-(x_{j-1} - x_j), -(x_{j-2} - x_{j-1}), \dots, -(x_1 - x_2)] = [x_j - x_{j-1}, x_{j-1} - x_{j-2}, \dots, x_2 - x_1]$ だから, これは $\alpha(x)$ を逆順に読んだリストであり, $\alpha((ir) \cdot x)$ と $\alpha(x)$ は $\gamma(\beta())$ をとると集合として一致する. したがって 4. の左辺 $= \gamma(\beta(\alpha((ir) \cdot x))) = \gamma(\beta(\alpha(x))) = IS(x) = E \circ (IS(x))$. \square

この定理は, 音列の領域での操作 k (可換図式の左側) が音程スケールの領域での操作 $f(k)$ (可換図式の右側) に対応することを示し, 音程スケールの操作は群 K と B 間の準同型写像 f によって単純化される関係にあることを示している. そして, i と r は音程スケールを反転させ, e と ir は音程スケールを変えず, 二つのグループに分かれることを意味する. さらに, もし音程スケールが inv において不変, つまり対称な形の音程スケールであるならば, これら 2つのグループは 1つになり, 4つの音列の形態において音程スケールが同じになることがわかる. これは, 対称な形の音程スケールは音列の形態を変えても音程スケールの「転調」を引き起こさず, 音程スケールがある意味で安定することを示している.

4.8 作曲への応用

以上のような音程スケールに関する理論を応用し, 実際に作曲した実例「Etude in Interval Scale {5, 6, 7}」を付録に掲載する.

本作品では, $\{5, \bar{6}, \bar{7}\}$ という音程スケールに基づいて旋律と和音を決定した. この音程スケールは, (1) 音程の転回について対称性を持つ, (2) 各音程が連続する, という 2条件を満たす音程スケールの中で要素数が最小のものである. 条件 (1) のもとでは, 定理 3 より, 音列の逆行型, 反行型, 逆行型の反行型の音程スケールが元の音列の音程スケールと同じになり, それらの操作の全てが音程スケールの「転調」なしに使用できる. また, この音程スケールは連続する音程を含むため, 定理 2 の系 2 より, この音程スケールに基づく音列中には 12 種類のピッチクラスが全て出現するという意味での無調性が確保される. そして要素数が最小のものを選択することで, 響きの偏りが大きくなると考えられる.

このような音程スケールを生成するには, 次のような列挙, スクリーニングの単純なアルゴリズムに従えば良い:

- \mathbb{Z}_{12} の全ての音程スケールを列挙する.
- inv で不変な音程スケールだけを残す.
- 連続な音程スケールだけを残す.
- 残ったものの中で要素数最小のものを残す.

以上のように音程スケールを選択した後, 本作品ではそれに基づく 4つの短音型 (旋律的音程のリストとして書くと, $[\bar{7}], [\bar{6}, \bar{5}], [\bar{7}, \bar{7}, \bar{7}], [\bar{6}, \bar{7}, \bar{5}, \bar{5}]$) をまず作り, その音型の順行, 逆

行, 反行, 逆行の反行, およびリズム長の拡大の操作を用いて楽曲を生成した. 八小節ごとの各セクションは, 基本音型と音型操作のうち, いくつかを選んだ組み合わせによって生成されており, 組み合わせの違いから様々なバリエーション (それぞれのバリエーションは二重縦線で区切られている) が生み出されている.

4.9 章のまとめ

本章では, 旋法の概念の変種であり, かつ音階の概念と双対の概念としての音程スケールという概念を提案した. そして音程スケールの概念が, n 音の調律システム \mathbb{Z}_n の部分群の生成系としての意義をもつことを示し, 生成される部分群が, 音程スケールの各要素の最大公約数 d で特徴付けられることを示した. また, 無調音楽や 12 音音楽への応用上重要な知識として, 音程スケールの生成する部分群がクロマティックになる条件を示した. その帰結として, Ligeti と Berg の例における音程スケールの選択が, 無調性と特徴的な響きという, 一見矛盾した特性に見えるものをうまく調和させる効果をもたらしているという解釈を導いた. これらの結果は, 音程スケールを作曲に応用する際, 音程スケールの選択を考える上で有用な手がかりとなるだろう. さらに, Berg の 12 音音列から導き出された音列の音程スケールを一般の音列に対して定義し, 音列に対する順行, 反行, 逆行, 反行の逆行という操作のなす群が, 音列の音程スケールの恒等写像と反転のなす群への群準同型によって単純化される関係にあることを示した. ここから, 音程スケールを不変に保ちながら 4 種類の音列操作の全てを使用できるためには音程スケールの転回に対する対称性を持つべきことが明らかとなった. 最後に, これらの考察結果を総動員して, 無調性, 響きの偏り, 音列操作における不変性をもつような音程スケールを生成し, 作曲に応用した.

今後の課題は, 音程スケールに基づくさらなる作曲と, 作曲支援技術の開発である. Berg のヴァイオリン協奏曲は「ある天使の思い出」に捧げられており, その背景には, Berg が可愛がっていたマノン・グロピウスの死があるといわれている. したがって, マノンの死への Berg の感情がこの曲に反映されていると考えられ, それが音程スケールにも関係していると考えるのは自然であろう. ここからは, 音程スケールがどのような感情を表現できるかという前章とも通底する疑問が現れる. この疑問は音程スケールに基づく作品をさらに作っていくことを通して徐々に明らかにしていきたい.

また音程スケールを作曲に応用する場合, ある音程スケールに基づく 12 音音列その他の音列が必要になるなどのケースが想定できる. しかし, そのような音列を発見したり, 存在を判定したりするのは困難であるため, 望ましい条件を満たす音列の存在判定法や, 存在する場合, 実際にコンピュータによって探索するアルゴリズムを開発することは有用だと考えられる. 実際, 12 音音列において $\bar{0}$ 以外の全ての音程をもつ総音程音列をコンピュータ・プログラムによって生成する研究は過去に行われてきた [35]. この問題を一般の音列および一般の音程スケールに拡張して捉え直すことで, 音程スケールを用いた作曲に活かすことは重要な研究課題の 1 つである.

また, Iannis Xenakis の作品 “Herma” においては集合論的な作曲方法が探求されたが,

音程スケールと集合論的な作曲方法を融合させることも今後の研究課題として考えられる。Xenakisはこの曲において集合論的な作曲法を試みており、ピアノの鍵盤からランダムに割り当てた A と B と C の三つのピッチ集合を用いて、それらの間のブール演算を用いて新しい集合を生成し、曲の形式を作り出している（図 4.6）。この方法を、ピッチの集合ではなく音程の集合である音程スケールとして読み替えることで、音程スケールを作曲に応用することができると考えられる。その際に、音程の集合はランダムに選ぶのではなく、より認知的な意味を持つような選択の仕方を考えることも重要であろう。

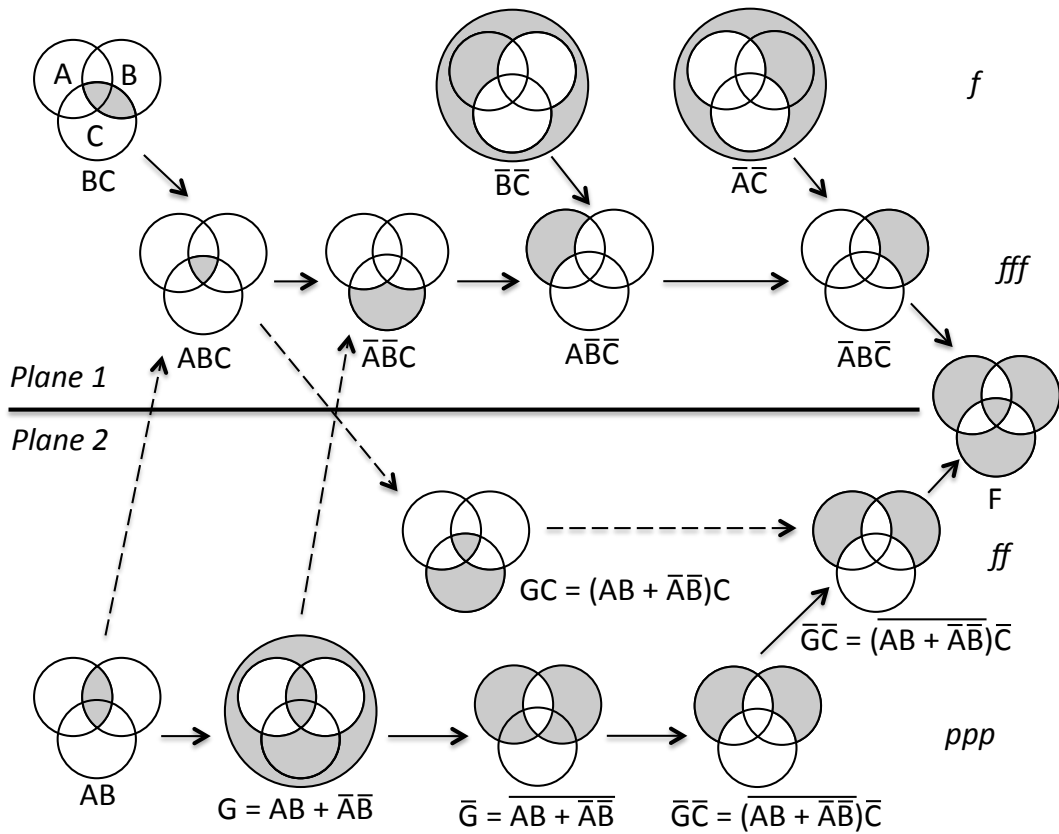


図 4.6. Xenakis が楽曲の形式を作るのに用いた集合論的な演算 [44].

第5章

サウンドファイルの対位法

5.1 音符からサウンドファイルへ

本章では、20世紀以降の音楽および情報技術の重要な動向をふまえ、その延長上に、近年の音楽情報環境の中で重要になってきていると考えられる音楽理論として「サウンドファイルの対位法」を提案する。

20世紀以降にはそれまでになく多様な実験的音楽や、工学の領域との融合した音楽研究が行われてきた。近年はインターネットの普及やコンピュータの性能の向上による情報環境の変化が起こっている。特に、デジタル信号処理技術のもたらしたインパクトは大きい。例えば、サンプリング音源の登場以前はライブであれ録音であれ、人間の演奏家が実際に演奏する必要があったが、サンプリング音源の登場により、コンピュータ上で楽譜を打ち込めば、録音された楽器の音で代用することで、かなりの程度のクオリティの演奏の再現が可能になった。近年では、最も再現が難しいと考えられる歌声の再現が「初音ミク」などのボーカロイドの登場により可能になってきた。また、作曲の領域では、ミュージック・コンクレートに端を発する録音素材を用いた音楽が盛んに創作されている。さらにインターネットの普及により、大量の楽曲データや音声データがやりとりされるようになってきている。

このような技術的な発展は、単に利便性や素材の新しさをもたらすだけではなく、音楽そのものの作られ方を大きく変質させる可能性をもっていると私は考える。例えば、一柳慧の「ピアノ・メディア」(1972)は、コンピュータの演奏するモーツァルトの楽曲を聴いた驚きをきっかけに作曲された曲であり[48]、人間らしさを追求するコンピュータ演奏とは対極的に、人間には演奏の難しい機械的な反復からなる楽曲をあえて人間が演奏するという、ある意味で異様な楽曲である。そこでは人間が演奏することの意味が問い直され、音楽のあり方や音楽に対する感性が同時代のテクノロジーのあり方によって変質してしまっていることを感じさせる。とはいえ、「ピアノ・メディア」は音符で書かれ、ピアニストが演奏するという従来の音楽の慣習の内部にある作品である。

では、より直接的に音楽が情報技術と結びつき、音楽が変容する可能性を突き詰めると何が生まれるだろうか。このような問題意識のもとで私が構想したのが「サウンドファイルの対位法」である。これは録音音源を素材として用いるミュージック・コンクレートの延長上で、サ

サウンドファイルを自動的に組み合わせるための音楽理論である。

そもそも対位法 [49, 50, 51] とは、複数の旋律を同時的に組み合わせ、良い響きを得るための技法であり、旋律の過度な跳躍を禁じるなどの旋律的な規則や、声部間の不協和音を解決するなどの声部間の規則からなる。歴史的には、対位法はパレストリーナやバッハをはじめとする、ルネサンス時代以来の多声音楽のスタイルに影響を与えてきた音楽理論である。対位法は禁則ルールの形で表せる部分が多くコンピュータ・プログラムで記述しやすいこともあり、コンピュータの登場後には、楽譜に基づく対位法的な音楽のさまざまな自動生成の手法が研究されてきた [15, 52, 53, 54, 55, 56, 16]。

他方で、本章で提案する「サウンドファイルの対位法」とは、従来、音符単位の音の組み合わせに関する音楽理論である対位法を、音符単位ではなく、サウンドファイル断片を単位素材とする組み合わせに拡張し、デジタルデータとしてサウンドファイルが氾濫する現代の音楽状況に対応しようとするものである。ここにおいては、音符の集合体としての楽曲はサウンドファイル断片のコーラージュ的な集合体へ、各声部の旋律はサウンドファイル断片の継起的な接続へ、声部間の関係は複数のサウンドファイルの同時的關係へと置き換えられる。そして、頭で音響をイメージし手書きで音符を書く作曲行為は、規則を満たす解の探索や条件によるサウンドファイルの検索といった自動処理に置き換えられる。

音符単位からサウンドファイル単位への移行において問題になるのは、第一に、サウンドファイルはデータとしては時間軸と音量との二次元のグラフ上の高速な振動にすぎず、それを眺めるだけではどのような聴覚的内容をもつかがわからないことである。そのため、人間はサウンドファイルを再生し、聴音や耳コピのような形で意味のある内容を抽出する作業を行うことになるが、人間がそのために費やす時間と労力には限界がある。ここで、コンピュータによってサウンドファイルの波形を音響分析し、聴覚的内容意味を抽出する処理が有用となる。第二に、音符は音高、音量、音価などにパラメータ化されており、楽器の出せる音の範囲の制約があるもののパラメータ空間内であらゆる音のバリエーションが自由自在に使用できるのに対し、サウンドファイルは、パラメータ化されていないオブジェクトであり、基本的にはそのままの形で素材とすることになる。したがって、サウンドファイルの対位法においては素材への依存度が高く、使用する素材に応じてどのように組み合わせるかをボトムアップに決定する必要がある。

本研究においてもサウンドファイルの対位法は古典的な対位法のように完成された体系には程遠く、現在進行形の研究対象であるが、本章では、現時点で考えられるサウンドファイルの対位法の一つの構成例を示す。ここでは、ピッチの比較的明瞭なサウンドファイルを主に想定し、ピッチ認識などの音響分析によって、素材となるサウンドファイルの特徴をデータとして記録し、そこからの検索によってサウンドファイル断片の良い組み合わせを見つける枠組みを記述する。それにより、従来は創作者が主観的に音を組み合わせていたミュージック・コンクレートやその延長上の音楽に対して一つの客観的な音の組織化方法を示し、また、通常の電子音楽の作曲における手探りの構成方法に対して、素材のデータからの検索・配置という自動化手法を導入した新たな作曲の実践のあり方を示したい。

次節からは、まずサウンドファイルの対位法の導入の背景となる関連の研究領域について

概説する。その後 5.6 節において具体的なサウンドファイルの対位法の構成法について記述する。

5.2 ミュージック・コンクレート

20 世紀の音楽史における一つの大きな転換点となる出来事として、1948 年のピエール・シェフェールによる、レコードやテープによる録音された自然音を素材として用いたミュージック・コンクレートの発明がある [57]。録音媒体に対する、切り貼りや重ね合わせや、逆向きの再生や再生スピードの変化といった電子的変調によって音楽を構成することが可能になり、音符に書かれたものを楽器で演奏するという音の出し方の制限を超えて、より多様な音素材を使用した実験的な音楽の創作が可能になった。1951 年にフランス放送局にはミュージック・コンクレートの探求グループが設置された。当時はそのような音楽を制作することは個人レベルでは困難で、専門的な設備のあるスタジオが制作拠点となっていたが、近年では録音音源が電子化し、テープという物理的な制約から開放されてより扱いが容易になり、パーソナル・コンピュータのメモリや計算速度の向上もあって、個人的な制作に録音音源を用いることは容易になっている。そして、録音素材は電子的な楽音生成技術とともに、電子音楽の重要な構成要素として用いられている。

5.3 サウンドアーカイブ

21 世紀において、録音素材をめぐる進展として重要な事柄として、Freesound[58] などに代表される Web 上のサウンドアーカイブの発達がある。映画やテレビ等の SE サウンドを含め、サウンドアーカイブの需要は以前から存在したが、有料であったり、得られる音や音楽のバリエーションが限られていた。しかし Freesound などのプロジェクトでは、ウェブサイトの上でフリーな素材として誰もが扱えるだけでなく、誰もがアップロードすることで素材のバリエーションの拡大に貢献することができる。今後もこのような動きが継続されれば、音楽の制作に必要な音素材がますます手に入りやすくなり、音楽の制作が豊かになると考えられる。

しかし、アクセス可能なアーカイブが充実しただけでは、自分で録音する労力は軽減されるものの、本質的に音楽に変化はもたらされるとは限らない。音楽そのものが本質的に変化するとすれば、それはアーカイブをいかに活用するかというレベルでの方法論が発明されなければならないだろう。また、アーカイブが拡大するに従って別の問題が発生する。それは、サウンドファイルが膨大になれば、サウンドファイルを聴くのにかかる時間が増大し、人間が聴くことのできる時間の限界を超えてしまうことでそれを活用するのが困難になるという問題である。また、必要なものがそうでないものの中に埋もれて区別することも難しくなる。そこで必要になるのが、検索の手がかりになる情報である。例えば Freesound 等のアーカイブにおいては、各サウンドファイルに対し、波の音、ピアノの音等のような、何の音なのかという情報や、どこで録音された音なのか、「Funny」、「Sad」などといったその音源の性格など、サウンドファイルの内容の把握の助けとなる複数のタグがつけられている。ユーザはそれを手がかり

に欲しい内容をもつサウンドファイルを検索することができる。

5.4 音楽情報検索

しかしここで問題がある。既存のタグの情報は言葉による性格づけであり、それは音声ファイルの特徴の記述としては粗雑な分類にすぎないものである。むしろ、時間変化する音のスペクトラム情報や発音タイミングなど、より詳細な情報がなければ、そのサウンドファイルの音楽的な固有性を捉えることは難しい。そこで、サウンドファイルをコンピュータによる様々な音響分析にかけることで内容を詳細に把握できるようにする必要がある。そのような音の情報を自動的に検出、検索する要求に応えるための研究領域として Music Information Retrieval (音楽情報検索) という工学的研究の分野がある。

例えば、コード進行の認識は音楽情報検索の重要な研究テーマの一つである。コードを認識することで、似たようなコード進行をもつ楽曲をアーカイブから検索し、楽曲を推薦するなどのことが可能になる。コード進行を認識するには、多数の音が重なっている音響データから音符データを抽出する多重ピッチ認識などの技術が必要になる。多数の音が重なっているとき、サウンドファイルは個々の音の波形が混ざり合って複雑な波形となる。その場合、元の波形がどのようなものであったかを復元することが多重に重なったピッチを認識する事であるが、楽譜のパート譜のように個々の音についての分離された情報がない状態から復元することは非常に困難な問題である。このように、様々な音が混ざり合った状態のサウンドファイルから、有用な情報を抽出することが音楽情報検索にとって重要であり、例えば、ピッチ、発音時刻をはじめとして、音色やサウンドファイル間の類似度など様々な情報を分析する研究が行われている。このような研究が多様化し、精度が向上することは、サウンドファイルのアーカイブを自動処理で使いこなして音楽を生成するための強力な基盤となると考えられる。

5.5 コンピュータと音楽理論

既に述べたように、サウンドアーカイブが充実し、音楽情報検索の技術が発展してきた現在、専門家でなくても自分のパソコンでそれらを活用できるようになりつつある。そのような基盤が整いつつある一方で、それをどのように使いこなし創作に取り入れることで新しいタイプの音楽を生み出せるかは、技術系の研究とは別の研究課題として残されている。それは情報技術の研究開発をキャッチアップしながら創作系の研究者や音楽理論家が行うべき研究だと考えられる。そして、技術的基盤と音楽の中身にかかわる研究を融合させることではじめて、本質的に新しい芸術音楽が生まれるのではないかと私は考える。

サウンドファイルの対位法と同様、既存の音楽の素材をもとに楽曲を生成するための先行研究としては、IRCAMが開発した即興演奏生成システム [59] がある。この研究は、既存の楽曲から抽出したファクター・オラクルと呼ばれる構造をもとに、もとの楽曲に出現する楽曲の部分をつなぎ合わせて楽曲を再生成するものである。また、既存楽曲の断片の再結合をもとに類似したスタイルの楽曲を作り出す David Cope [60] や、類似した楽曲の文法を抽出し、そこ

から楽曲を生成する T. Kohonen [14] の手法も、似たようなタイプの楽曲生成研究である。

しかし、これらの楽曲生成手法は、既存の楽曲のスタイルを模倣することが目指されており、既存楽曲と無関係な楽曲を生成することは困難である。むしろ本研究にとって必要なのは、互いに無関係な複数のサウンドファイルから断片を取り出して組み合わせることで、新たな組み合わせのもとで楽曲を作る方法論である。なぜならサウンドアーカイブは、場所や状況や発音体、録音した人などがばらばらで、互いに無関係なサウンドファイルがとにかく集められた集合体であり、元の曲が無い状態から素材をコラージュして、新たな音楽スタイルにまとめあげるような手法が必要になるからである。そのために役立つと思われるのは、音楽的内容をコンピュータで記述するための形式的基盤を提供するような音楽理論である。

本節では、参考となる既存研究として、(1) 楽曲のアルゴリズムによる再構成、(2) 数学的な音楽理論、(3) compositional design という三つの研究領域を参照し、サウンドファイルの対位法との接点を探る。

5.5.1 楽曲のアルゴリズムによる再構成

西洋音楽において19世紀後半から20世紀初頭には、古典的な機能和声飽和し、そこから逸脱した音楽の創作が試みられるようになった。シェーンベルクらの12音音列に基づく音楽や、ブーレーズらによるセリー音楽、クセナキスらによる確率論的手法の導入、コンピュータによるアルゴリズム作曲等、伝統的な音楽理論の代わりに数理的な手法が導入される流れが生まれた。

そのような音楽においては、古典的な音楽とは異なり、作品の作り方そのものがシステムティックであり、コンピュータ・プログラムを書くことと作曲することが等価に近づいたと考えられる（それらの作品が必ずしもコンピュータを用いて自動的に作られたというわけではないが）。そして一旦楽曲がコンピュータ・プログラムとして再構成されると、パラメータを変えてその楽曲をモデルとした新しい楽曲のインスタンスがいくらかでも生成できるようになる。

このような方向性で生成アルゴリズムが再構成された例としては次のようなものがある。György Ligeti の「ピアノのためのエチュード第1番」は部分的に Pd のプログラムとして再構成されており [61]、Iannis Xenakis の “Herma” [62]、Pierre Boulez の “Structures Ia” [63]、James Tenney の “Spectral CANON For CONLON Nancarrow” [64] は Open Music 等のプログラミング・ツールを用いて再構成されている。またライヒの「ピアノ・フェーズ」は構造が明晰であり、コンピュータ・アルゴリズムに還元することは比較的容易である。このような現代のシステムティックな楽曲をプログラムとして再構成することで、自動作曲のための作曲モデルとすることは1つの有力な方法である。

しかし、このような方法の欠点は、モデルとなる一つの楽曲のスタイルから離れるのが難しく、多様な楽曲を生成するのは困難だと考えられることである。したがって、より汎用的な音楽スタイルを捉えられるような枠組みが望まれる。

5.5.2 数学的な音楽理論

20 世紀においてシステマティックな手法で作曲を行う作曲家が現れた一方、理論家の研究においてもシステマティックかつ数学的な音楽理論が登場した。その原動力となったのは、古典的な機能と声による分析が通用しない楽曲に対する新しい分析手法の必要性であり、ピッチクラス集合論 [65]、変換理論 (Transformation Theory) [45]、ネオリーマン理論 (Neo-Riemannian Theory) [66] などの数学的な音楽理論が発展してきた。それ以前にも、数学と音楽の関係はあったものの、それは文献 [67, 68] のような、調律や協和性などの音の物理学に関連するタイプの研究ものが主であり、曲の内容を分析するようなものではなかった。しかし 12 音音楽などの影響で、音列の操作を群論的に記述するなど、音楽的なオブジェクトとその変換操作を定義し、分析するような手法が見いだされ、音楽の内容を数学的に記述することができるようになってきている [45]。また音楽的なオブジェクトにおける類似性を計算する様々な手法も考案されてきた [69, 70]。また楽曲分析プログラムに応用された例もある [71]。しかし、このような数学的な音楽理論は、音楽の生成のためではなく既存の楽曲の分析を主な関心として発展してきたため、コンピュータ・プログラムとして実装されたり音楽の生成に用いられるような応用はあまり多くはなされていない。だが、このような理論はコンピュータとも親和性が高いため、音楽の生成と関連しながら発展する可能性は高いと考えられる。

例えば、ネオリーマン理論は和声法を数学的に構成する音楽理論であり、三和音を使用しつつも、古典的な調性音楽から逸脱した楽曲をも古典的な調性音楽と共に統一的に扱う必要性から生まれた理論である。その主要文献の一つ “Neo-Riemannian Operation and Their Tonnetz Representations” [66] においては、音楽理論において長短の三和音が基本的な和声進行となる P (Parallel), L (Leading-tone exchange), R (Relative) のもとで、構成音の動きが音の動きが最も少ないという意味で特別に良い和音の形であることを示している。この観点から、24 平均律などの 12 音ではない調律システムにおいて同様の計算を行う事で、その調律システムの基本和音をコンピュータで計算して導き出し、24 音の調律システムのための和声法を生成するなどといった形で、音楽の生成の研究に応用するといったことも考えられる。

このように、数学的な音楽理論は、汎用性が高くコンピュータへの実装もしやすい音楽的形式を提供するという意味で、サウンドファイルの対位法が必要とするものに近いと考えられる。

5.5.3 Compositional Design

数学的な音楽理論の概念の作曲への応用を考えると、数学的な音楽理論の研究の流れに属しつつも、作曲家による作曲のための方法論として Robert Morris の構想した compositional design の考え方 [47] は本研究にとって示唆的な部分を含む。compositional design とは主に音列を垂直に組み合わせた二次元の配列のことであり、それをガイドとしながら、比較的自由にそれを解釈して実際の音楽をリアライズする枠組みである。図 5.1 に compositional design

X	7 3 9 2 5 4	$T_3 X$	T 6 0 5 8 7
X	7 3 9 2 5 4	$T_6 X$	1 9 3 8 E T
X	7 3 9 2 5 4	$T_9 X$	4 0 6 E 2 1
X	7 3 9 2 5 4	X	7 3 9 2 5 4

図 5.1. Compositional design の例 [47]. 左図は音列 $X = [7, 3, 9, 2, 5, 4]$ をカノン状に配置したもの. 右図は X を i 半音上へトランスポートした $T_i X$ を用いたもの. ただし T はピッチクラス 10 を, E はピッチクラス 11 を表す.

の例を示す:

Morris は, compositional design を楽曲の作曲前のスケッチのようなものであるが, より正確には通奏低音のようなものであると述べている. つまり, compositional design によってピッチクラスは定まったとしても, 音域やリズムといった不確定要素が存在し, それをリアライズする際に補うが必要になる. そこにおいて不確定要素の扱い方の自由度が高いため, 柔軟に作曲が行えるというわけである.

このシステムを本研究に応用することを考えた場合, compositional design をガイドとして, それにマッチするようなサウンドファイルを音響分析の結果をもとに検索して音符の代わりにあてがうことで, 不確定要素のリアライズをサウンドファイルによって行うということが考えられる. これによって, 音楽理論とサウンドファイルを用いた作曲の間に橋をわたすことができる.

ただし, この手法をそのまま実装するには困難がある. それは, compositional design に実際にマッチするようなサウンドファイルが存在しない可能性が高いという点である. 音列は順序づけられたピッチクラスの集合であり, あらかじめ与えられた音列の順序とサウンドファイルのもつピッチクラスの出現順序が一致することが稀であるため, 欲しい音列を検索してもヒットするとは限らない.

そこで次節では, compositional design を参考にしつつも, よりサウンドファイルの検索によるリアライズのガイドとしてより適した枠組みを考える.

5.6 サウンドファイルの対位法の構成

5.6.1 ピッチクラス集合への包含によるリアライズ

サウンドファイルの組み合わせを compositional design のもとでリアライズしようとする際に問題となるのは, compositional design の音列に実際にマッチするようなサウンドファイルが存在しない可能性が高いという点であった. そこで, この制限を緩和するために, 音列の順序づけを外し, 順序づけられていない集合とそれに対する包含に着目する. なぜなら, 順序

のないピッチクラスの集合の場合、ピッチクラスの出現順序の制限や単旋律という制限がなく
なり、該当するサウンドファイル断片が検索にヒットする可能性が高くなるからである。図
5.2 に順序づけられていない集合によるサウンドファイルのリアライズのしかたを示す。

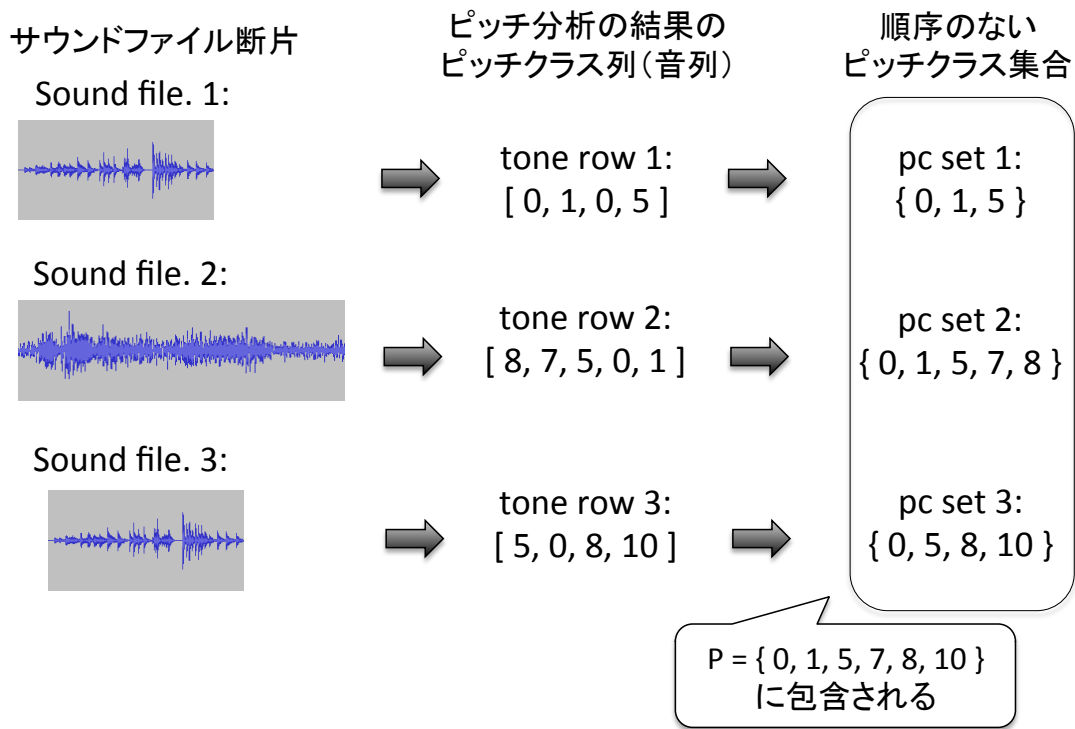


図 5.2. ピッチクラスへの包含関係によるサウンドファイルのリアライズ。図左の各サウンドファイルのピッチ分析によって、対応するピッチクラスの列を得る (図中央)。各ピッチクラスの列に出現する要素を集めたピッチクラス集合を作る (図右)。目的のピッチクラス集合 P に包含されている (pc set 1 $\in P$ かつ pc set 2 $\in P$ かつ pc set 3 $\in P$ である) ため、図の3つのサウンドファイルは P によってリアライズが可能だと考える (3つのサウンドファイルは同時再生が可能である)。

この方法により、リアライズされたサウンドファイルの響きを、目的のピッチクラス集合 P によって緩やかに統制することができる。

5.6.2 ピッチクラス集合間の遷移ネットワーク

次に問題になるのは、あるピッチクラスから別のピッチクラスへの音楽的な意味のある遷移の仕方である。そこで、ピッチクラスの集合とその遷移を扱う理論において重要な概念である“voice-leading persimony”の概念を導入する。voice-leading persimonyとは、あるピッチクラスから別のピッチクラスへの遷移において、ピッチクラスが保続されるものが多く、かつ、動きのあるピッチクラスにおいても、動きができるだけ小さい方が良いという基準である。

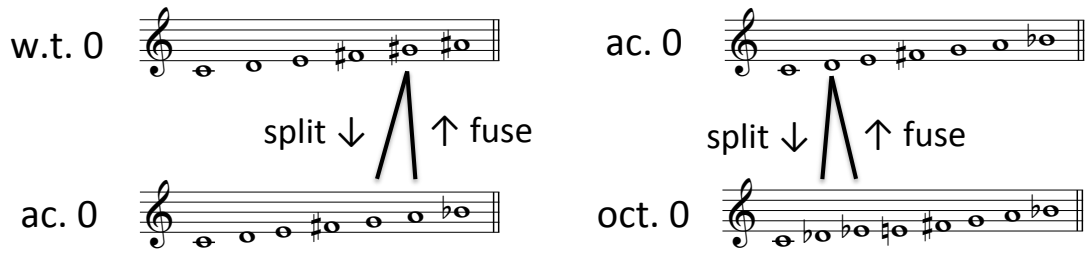


図 5.3. split と fuse. ただし, w.t. 0 は C のピッチクラスから始まる全音音階 (whole-tone scale) のピッチクラス集合を表し, ac. 0 は C のピッチクラスから始まる acoustic scale のピッチクラス集合を表し, oct. 0 は C のピッチクラスから始まる octatonic scale のピッチクラス集合を表す [72].

(Persimony は日本語で儉約を意味する). つまり, どれだけピッチクラスが滑らかに遷移したかという基準である.

Clifton Callender は論文 “Voice-leading parsimony in the music of Alexander Scriabin” [72] の中で, 作曲家 Scriabin が好んで使用した複数のピッチクラス集合の関係を論じており, voice-leading parsimony を実現するような数種類のピッチクラス間の演算を示している. ここでは, split と fuse という一对の演算のタイプが定義されており, 実際に Scriabin の作品の例を分析しながら, そのような演算によって楽曲のピッチクラスの推移の法則性を記述できることを示した. split とは, ある一つのピッチクラスがそれに隣接する二つのピッチクラスに分岐する演算であり, fuse は逆に, 全音離れている二つのピッチクラスを, それらの間にある一つのピッチクラスへ統合する演算を指す. 図 5.3 に, split と fuse の実例を示す.

ここで, acoustic scale には 12 種類の移調形があるが, octatonic scale には 3 種類, whole-tone scale には 2 種類の移調形が存在する (octatonic scale および whole-tone scale は作曲家 Messiaen のいう「移調の限られた旋法」である). 図 5.4 にこれらの移調型を示す.

Callender はそれらのピッチクラス集合の間の split と fuse による遷移可能性を表すネットワーク (図 5.5) を示している.

このような, あるタイプの演算に基づくネットワーク中の遷移によって compositional design のような潜在的な構造を作ること, 数理的に規定された楽曲のスタイルが生成できると考えられる. そして, compositional design におけるような, 実際にピッチクラス集合に該当するサウンドファイルが存在するかどうかという問題も, 順序づけられた音列ではなく, 順序付けられていない集合を用いることで, 解消される可能性が高められると考えられる.

なお, ここで挙げたピッチクラス集合のネットワークは一つの特例であり, どのようなピッチクラス集合やその推移規則を用いるかは様々な可能性がある. その中で, 素材として用いるサウンドファイルに適したピッチクラス集合やその推移規則と, そうでないものが存在すると考えられる. もし適さないネットワークを構築した場合リアライズする素材が見つからないことも起こりうる. したがって素材とネットワークは相互依存的に決定する事が望ましい.

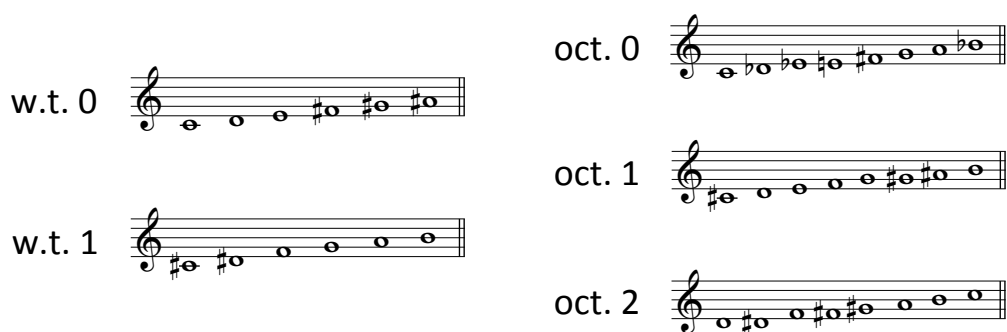


図 5.4. whole-tone scale の二種類の移調形, および octatonic scale の三種類の移調形. ピッチクラス集合の名前につく数字は元型からのトランスポジションの度数 (半音の数) を表す. w.t. 3 は w.t. 0 に集合として一致し, oct. 3 は oct. 0 に集合として一致するため, 12 種類よりも移調形の個数が少なくなる [72].

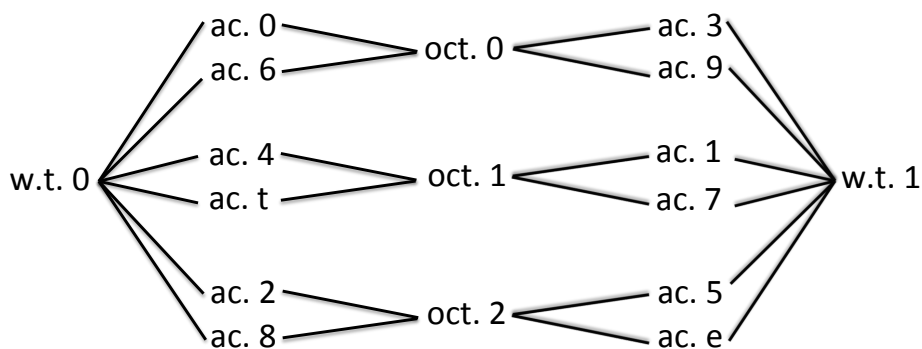


図 5.5. whole-tone scale, acoustic scale, octatonic scale の間の split と fuse による遷移のネットワーク. 線を辿ってピッチクラス集合を遷移することで, voice-leading persimony の意味での安定性が保証される [72].

5.6.3 声部の処理

前節で述べたような、楽曲のスタイルを構造的に規定するピッチクラス集合間のネットワークを用いて、そこに包含されるサウンドファイルを複数トラック並列する形でのリアライズを考慮することができる。このような枠組みが「サウンドファイルの対位法」の構成法の一例である。図 5.6 にその概念図を示す。

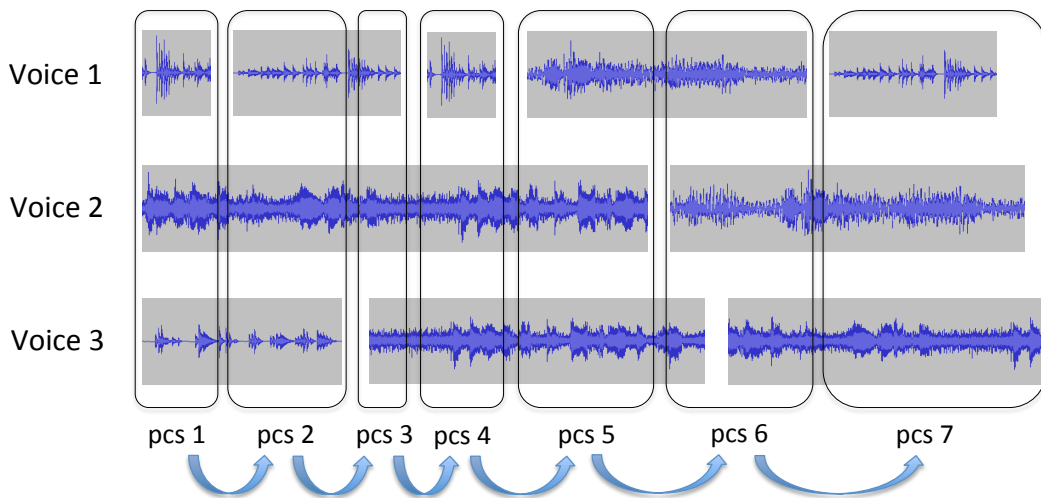


図 5.6. サウンドファイルの対位法の概念図（三声の場合）。ピッチクラス集合間のネットワークにそって遷移する pcs（pitch class set）に従って、常に三つのサウンドファイルがリアライズされている。ただし、横方向はサウンドファイルの継ぎ足しによる「声部」の時間的経過を表し、縦方向は各声部のサウンドファイル同士の同時的な重なりを表す。

この図において、サウンドファイル間の同時的な関係性は、共通するピッチクラス集合に包含され、ある一つの声部のサウンドファイル断片が終了すると、そこに新たなサウンドファイルが継ぎ足され、そのサウンドファイルのもつピッチクラス集合（音響分析によって得られたもの）次第でピッチクラス集合の遷移が起こる。ただし、その遷移は、どのようなサウンドファイルでも許されるのではなく、split と fuse のような、ある決まった演算によって遷移する。このような潜在的なピッチクラス集合のネットワークによって楽曲スタイルが規定される。

5.6.4 発音位置の特定と同期

ピッチクラス集合によるコントロールだけでなく、リズムの制御をすることも重要である。そのために、例えばサウンドファイル内の発音位置を特定し、図 5.7 のように声部間で同期をとるような対位法を構成することも考えられる。

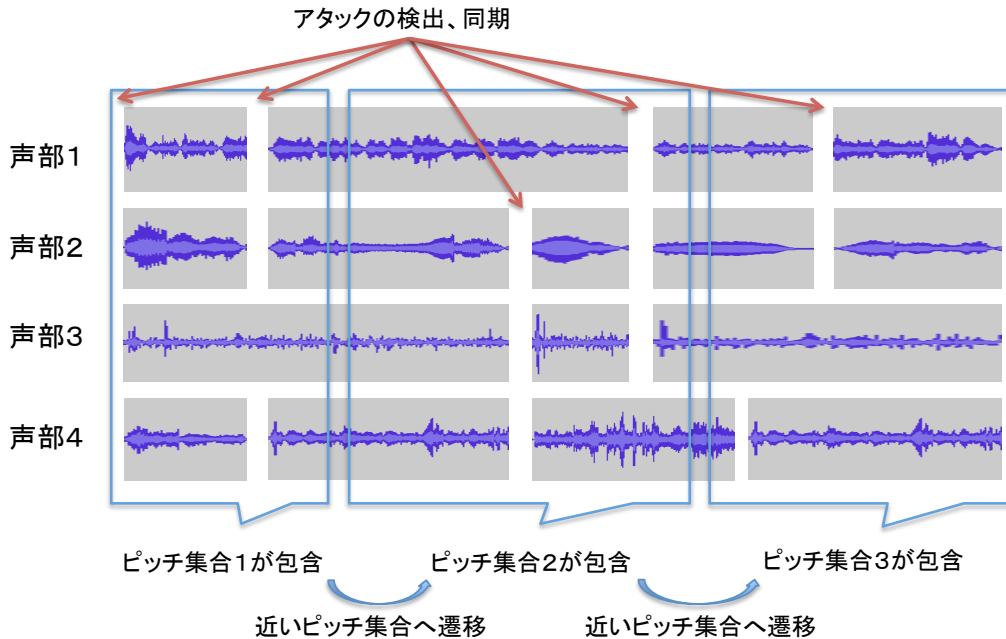


図 5.7. リズムのコントロールを含むサウンドファイルの対位法の概念図. ピッチクラス集合への包含関係に加えて、発音位置を検出しサウンドファイル間の同期をとる.

5.6.5 使用するサウンドファイルと特徴量

サウンドファイルの対位法においては、ピッチや発音位置の特定以外にも様々な音響分析を用いることが考えられる。例えば音響的な特徴量を分析する際には、音響分析のためのプログラミング・ツール ESSENTIA[73]などが開発されており、最先端の研究の成果が比較的容易に使用できる場合もある。ESSENTIAにおいて分析できるスカラー値の特徴量としては表 5.1 のようなものが挙げられる。

このような特徴量の使用方法としては、用意したサウンドファイル群に対して、ある特徴量に関して音響分析し、ランキングの上位のもののみを用いて楽曲をリアライズすることにより、楽曲に特徴を与えるという方法が考えられる。さらに高度な特徴量の活用法としては、これらの特徴量をベクトルとして入力し、多重回帰分析などの機械学習 [74] を行い、ある形容詞への当てはまりの良さを実現するような特徴量ベクトルを学習することで、求める質のサウンドファイルを検索できるようにする等の方法も考えられる。それが可能になれば、求める質を持った楽曲を作り出せるようになるだろう。

5.7 章のまとめ

本章では、電子音楽の時代に対応した音楽理論の必要性から私が案出した「サウンドファイルの対位法」という概念を提案した。それは通常の音符を単位とした対位法をサウンドファイ

特徴量	内容
duration	サウンドファイルの長さ
effectiveDuration	サウンドファイルの中で知覚的に有意な部分の長さ
chordsChangesRate	和音の変化率
key_strength	調性の強さの度合い
loudness	音量の大きさ
beatsLoudness	全周波数帯域におけるビートの強さ
loudnessBandRatio	ある周波数帯域におけるビートの強さ
bpm	テンポの速さ
tuningFrequency	チューニング周波数
energy	エネルギー
flux	スペクトルの変化の大きさ
zeroCrossingRate	音響信号の符号の変化の速さ
oddtievenharmonicenergyratio	偶数倍音と奇数倍音成分の比率
spectralComplexity	スペクトルの複雑性
flatness	スペクトルの平坦性
inharmoniccity	不協和性
dissonance	感覚的な不協和性
centroid	スペクトルの重心
variance	スペクトルの分散
kurtosis	スペクトルの尖度
skewness	スペクトルの歪度
danceability	ダンスへの適合性

表 5.1. ESSENTIA において計算できる様々な音響的特徴量 (抜粋).

ルを単位素材として扱うために拡張したものであり、使用する素材としてのサウンドファイルの特徴を音響分析して記録し、その情報をもとに良い音の組み合わせを求める枠組みである。具体的な方法論として、ピッチクラス集合のネットワークへの包含によるサウンドファイル断片のリアライズや、発音位置の推定によるサウンドファイル間の同期、音響的特徴量を元にしたサウンドファイルの選択を提案した。これによって、従来は創作者が主観的に音を組み合わせていたミュージック・コンクレートに対して、1つの客観的な音の組織化の方法を示し、また、通常の手探りの作曲行為に対して素材の探索や配置の自動化を電子音楽の作曲に導入することを提案した。

私はこのような作曲の方法論が今後ますます発展し、ポピュラリティを獲得するような時がそう遠くない将来に来る可能性があると考えている。例えば、サウンドアーカイブのウェブサイトにサウンドファイルの対位法の構成アルゴリズムが連動して実装されているシステムを想像してみる。そこにおいては、アーカイブの各サウンドファイルとそれに対して予め音響分析したデータがセットで利用できるようになっており、ユーザが自分で音響分析プログラムを作らなくても分析情報を作曲に活用できるようなインターフェースが整備されているという状況を想像するのはそれほど困難ではないだろう。そのようなシステムを現実のものとするには、音楽情報検索をはじめとして、電子音楽の創作論、数理的な音楽理論、インターフェース研究などの領域を横断的する研究をさらに進めていくことが必要である。

第6章

結論

最後に、本論文を総括する。

第2章では、あたらしい旋律スタイルの生成を目的とした形式文法的な旋律モデルを構築し、音型の形成ルールの整合的な集合を発見する問題として生成手法を定式化した。そこでは、作曲家 Messiaen による、音型の引用と合成によって旋律スタイルを作るプロセスを参照・応用し、また多声音楽にも対応するため、声部ごとの旋律法を同時共存した形での楽曲の生成を可能にするモデル化を行った。そして、クラシファイア・システムに基づくルール生成手法を応用することで、旋律法の創発を実現した。しかし、既存のスタイルの模倣ではなく新しいスタイルを生成する本研究では、どのように妥当性を評価するかが難しい問題である。そこで私は作曲の専門家の聴取による評価が有効と考え、ルール数および声部間での旋律法の共有／非共有という二つのパラメータを設定して評価実験を行った。結果として、ルール数を小さく抑えた場合に、旋律スタイルの独創性や統一性などの様々な観点からスタイルの評価が向上することや、声部間の複数の旋律法の共存の効果が示され、生成される旋律スタイルの質をパラメータでコントロールできるようになった。一般に自動作曲の研究においては、楽曲の評価をどのように行うかが難しく、生成手法の記述が主で、評価実験は行わなかったり簡単な評価実験で済ませてしまう場合も多い。しかし本研究では評価実験を重視して生成モデルを作り、評価値に対して因子間の交互作用を含む統計分析を行った。この点においても、自動作曲の研究水準の向上に寄与できたと考える。

第3章では、特定の感情を表現する旋法を生成する手法を提案した。旋法と感情の対応関係は古代ギリシャ時代からの問題であるが、旋法の種類は無数にあるため、対応関係に対して心理実験による客観的な裏付けを与えることが困難であった。そこで私は発想を転換し、強化学習を用いた機械学習的アプローチをとり、人間のもつ特定の感情に対して旋法の方を適応的に変化させ、感情に適合した旋法を得る手法を開発した。実際、特定の旋法が使用されるようになる経緯は、長期間にわたる動的、適応的な過程だと考えられる。本研究はそのような適応的過程を短時間でシミュレートでき、古くから十分には解決できないでいた問題に解決を与える可能性があるという意味で高い価値があると思われる。実験結果としては、4つの種類の感情に対して感情の表現力をもつ旋法を獲得することができた。またそこでは、喜びの旋法としてミクソリディア旋法が生成されたり、喜びと悲しみの旋法として日本の伝統的な旋法が埋め込ま

れた旋法が生成されるなど、既存の旋法の意義の再発見や学習における文化的な影響を暗示するかのような興味深い結果も出ている、今後はより詳細な実験を行うとともに、このモデルを他の音楽的要素にも拡張したいと考えている。本手法は以下のような応用上の大きな可能性を秘めている。(1) 音楽学や音楽心理学への構成論的 (synthetic) なパラダイムの導入：構成論的手法とは、実物 (正解) となる楽曲やパターンのサンプルをまず生成した上で、その分析から何かを解明するというような、従来とは逆の手順をとる手法である。それにより、例えば文献 [75] のような、旋律パターンとその表現内容の主観的な対応づけに客観的な裏付けを与えるといった研究につながる可能性がある。(2) 微分音による調律システムなどの人間の作曲者が踏み込みにくい領域の開拓。(3) 楽曲への感情や雰囲気との付与：動画や劇に付随する音楽やカフェのBGMなど、様々なシチュエーションに対してふさわしい楽曲が欲しいといったニーズに応えるには、感情や雰囲気を付与する技術は極めて重要であろう。(4) システムのユーザ自身の感情をもとにした楽曲生成：本手法においては、単にできあがったものを提供するのではなく、ユーザが自分自身で音楽理論へと感情を付与できる点で、ユーザが表現の主体となれるという利点がある。また、本手法は既存の旋法テンプレートから選択するのとは異なり、新しく旋法を生成するものであるためクリシェに陥らずにすむ可能性が高いという利点もある。(5) 音楽療法への応用：クライアントの感情に働きかける音楽を生成することは音楽療法にとって重要であり、本手法はそのために応用できる可能性がある。

第4章では、第3章の研究の副産物として、音階 (scale) という極めて基本的な概念と双対な「音程スケール (interval scale)」という音楽学的な概念を考案 (発見) した。そしてこの概念の適用例と解釈できる楽曲として Ligeti と Berg の楽曲を挙げて分析を行い、この概念が作曲に効果的に用いられることを例証した。また代数学的な考察から、この概念が無調性に結びつきやすいことを示し、この概念を作曲に応用する際に役立つ数学的な命題を見いだした。そして、実際に音程スケールを生成し、それに基づいて作曲を行った。数学的な音楽理論はともすれば実践に結びつかない机の上だけのものになりがちな面があるが、本研究では、作曲に実際に活用するために数学的命題を導くという実践的に意義のある研究ができたのではないかと考える。ただし音程スケールに基づく作曲例はまだ少ない。今後も作曲を継続する中でその可能性と限界をより明確にしていきたい。

第5章では、従来、音符単位の組み合わせに関する音楽理論である対位法を、電子音楽に対応するためにサウンドファイル単位素材とする組み合わせに拡張したものと、サウンドファイルのための対位法という音楽理論を提案した。それにより、従来は創作者が主観的に音を組み合わせていたミュージック・コンクレートに対して一つの客観的な音の組織化方法を示した。そして、通常の作曲における手探りの作曲プロセスに対して、素材の検索という自動化を電子音楽の作曲に取り入れた新たな作曲の実践の形を示した。サウンドファイルはその内容を音響分析しないと聴覚的内容を検出する事ができないが、音響分析を高精度で行うことは専門家にとっても困難な研究課題であり、最も基本的なピッチ認識の問題ですらいまなお研究されている。そうした基礎技術の精度が十分でなければ、音響分析の結果を利用するサウンドファイルの対位法も精度が上がらない。したがって、サウンドファイルの対位法はそうした研究領域の進展とともに今後も発展する余地を残している。

本論文では、楽曲そのものではなくその背後に横たわる音楽理論を生成の対象とすることによって新しい音楽スタイルを作り出すという研究目的を掲げた。この目的に対して私は、各章においてそれぞれの音楽理論の領域で適した手法を考案したり実際に音楽スタイルを作り出し、一定の成功をみたと考える。そして、コンピュータや数理的な手法が自動作曲における創造性に寄与しうることを示せたのではないかと考える。また、本研究は東京芸術大学大学院の先端芸術表現専攻の博士課程での研究として、芸術と科学技術の間の領域を横断的に行うという計画のもとで行われた。その中で、音楽の創作や理論の研究が音楽の研究領域だけで自己完結するのではなく、機械学習、音楽心理学、数学、音響信号処理、音楽情報検索といった他の研究領域と相互に関わる研究を行うことによって、新しい視界が開け、新しい芸術の形が生まれうることを示せたのではないかと考える。

参考文献

- [1] O. Messiaen: *Technique de mon langage musical*, Alphonse Leduc (1944). Satterfield, J. (Translation): *The Technique of My Musical Language*, Alphonse Leduc, 1956.
- [2] Lerdahl, F., and Jackendoff, R.: *A Generative Theory of Tonal Music*, MIT Press, 1983.
- [3] G. Nierhaus: *Algorithmic Composition: Paradigms of Automated Music Generation*, Springer, 2009.
- [4] P. Prusinkiewicz: Score generation with L-systems, Proc. *ICMC*, pp.455–457, 1986.
- [5] P. Worth and S. Stepney: Growing music: musical interpretations of L-systems, Proc. *EuroGP*, pp.545-550, 2005.
- [6] R. Stocker, H. Jelinck, B. Burnota and T. Bossomaier(Eds.): *Complex Systems: From Local Interactions to Global Phenomena*, J. McCormack: Grammar-Based Music Composition, pp.321-336, IOS Press, 1996.
- [7] B.F. Lourenco, J.C.L. Ralha and M.C.P. Brandao: L-Systems, Scores, and Evolutionary Techniques, Proc. *SMC*, pp.113-118, 2009.
- [8] R.L. DuBois: Applications of generative string-substitution systems in computer music, PhD thesis, Columbia University, 2003.
- [9] D.A. Ashlock, S.P. Gent and K.M. Bryden: Evolution of L-systems for compact virtual landscape generation, Proc. *CEC* Vol. 3, pp.2760-2767, IEEE Press, 2005.
- [10] P. Mueller, P. Wonka, S. Haegler, et al.: Procedural Modeling of Buildings, *ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH 2006* Vol.25, No.3, pp.614–623 (online), DOI: 10.1145/1141911.1141931, 2006.
- [11] 平田圭二, 青柳龍也: バービーブーン: 音符レベルでユーザ意図を把握して編曲を行う事例ベースシステム, *情報処理学会研究報告*, Vol.2000, No.94, pp.17-23, 2000.
- [12] 平田圭二, 青柳龍也: 音楽理論 GTTM に基づく多声音楽の表現手法と基本演算, *情報処理学会論文誌*, Vol.43, No.2, pp.277-286, 2002.
- [13] J. Gillick, K. Tang, and R. Keller: Learning Jazz Grammars, Proc. *SMC*, pp.125-130, 2009.
- [14] T. Kohonen: A Self-Learning Musical Grammar, or “Associative Memory of the Second Kind,” Proc. *IJCNN*, pp.1-5, 1989.

- [15] M.V. Mathews and J.R. Pierce (Eds.): *Current Directions in Computer Music Research*, Schottstaedt, W.: Automatic Counterpoint, pp199-214, MIT Press, 1989.
- [16] T. Tanaka, T. Nishimoto, N. Ono and S. Sagayama: Automatic Music Composition Based on Counterpoint and Imitation Using Stochastic Models, Proc. *SMC*, pp.330-337, 2010.
- [17] J.H. Holland: *Adaptation in Natural and Artificial Systems*, pp.171-198, MIT press, 1994.
- [18] 電気学会 GA・ニューロを用いた学習とその応用調査専門委員会 (編): 学習とそのアルゴリズム, 玉置久: 遺伝アルゴリズムと機械学習の基礎, pp.86-101, 森北出版株式会社, 2002.
- [19] T. Taura, I. Nagasaka and A. Yamagishi: Application of evolutionary programming to shape design, *CAD*, Vol.30, No.1, pp.29-35, 1998.
- [20] プラトン: 国家 (上) (藤沢令夫訳), pp.209-213, 岩波書店, 2005.
- [21] アリストテレス: 政治学 (山本光雄訳), pp.368-381, 岩波書店, 1997.
- [22] M. P. Kastner and R. G. Crowder: Perception of the major/minor distinction: IV. Emotional connotations in young children. *Music Perception*, 8, 189-202, 1990.
- [23] D. S. Hill, S. B. Kamenetsky, and S. E. Trehub: Relations among text, mode, and medium: Historical and empirical perspective, *Music Perception*, 14, 3-21, 1996.
- [24] D. Ramos, J. L. O. Bueno and E. Bigand: Manipulating Greek musical modes and tempo affects perceived musical emotion in musicians and nonmusicians, *BRAZILIAN JOURNAL OF MEDICAL AND BIOLOGICAL RESEARCH*, Vol.44(2), pp.165-172, 2011.
- [25] A. P. Merriam: *The anthropology of music*. Northwestern University Press, 1964.
- [26] S. Sadie et al. (Eds.): *The New GROVE Dictionary of Music and Musicians*, Macmillan Publishers Limited, London, 1980.
- [27] R. S. Sutton and A. G. Barto: *Reinforcement Learning: An Introduction*, The MIT Press, 1998.
- [28] S. Le Groux, and Paul F. M. J. Verschure: Towards Adaptive Music Generation by Reinforcement Learning of Musical Tension, Proc. *SMC*, pp.160-165, 2010.
- [29] M. V. Butz, O. Sigaud, G. Pezzulo and G. Baldassarre (Eds.): *Anticipatory Behavior in Adaptive Learning Systems*: A. Cont, S. Dubnov, and G. Assayag: Anticipatory Model of Musical Style Imitation using Collaborative and Competitive Reinforcement Learning, pp.285-306, Springer, 2007.
- [30] P. N. Juslin and J. A. Sloboda (Eds.): *Handbook of Music and Emotion - Theory, Research, Applications*, Oxford University Press, 2010.
- [31] John Rahn, *Basic Atonal Theory*, SCHIRMER BOOKS, 1980.
- [32] O. Messiaen: *Mode de valeurs et d'intensités : pour piano*, Durand, 1950.
- [33] R. Johnson: *Messiaen*, pp.105-107, Omnibus Press, 2008.

- [34] G. Ligeti: *Études pour piano premier livre*, pp.14-19, SCHOTT, 1986.
- [35] 柴田 南雄: 音楽の骸骨の話: 日本民謡と 12 音音楽の理論, pp.93-171, 音楽之友社, 1978.
- [36] 小鍛冶 邦隆: 作曲の技法: バッハからヴェーベルンまで, pp.107-122, 音楽之友社, 2008.
- [37] A. Berg: *Konzert für Violine und Orchester*, pp.219-241, Breikopf & Härtel, 2010.
- [38] Douglas Jarman: *The Music of ALBAN BERG*, pp.136-137, University of California Press, 1979.
- [39] L. Bigo, J.L. Giavitto and A. Spicher: Building Topological Spaces for Musical Objects, Proc. *MCM*, pp.13-27, Springer, 2011.
- [40] 三宅 敏恒: 入門代数学, pp.1-35, 培風館, 1999.
- [41] 田村 孝行: 半群論, pp.1-12, 共立出版, 1972.
- [42] J. Brank, C. Hailey, and D. Harris (Eds.): *The Berg-Schoenberg Correspondence: Selected Letters*, W W Norton & Co Inc, 1987.
- [43] J.Rose: *A Course on Group Theory*, pp.50, DOVER, 2012.
- [44] I. Xenakis: *FORMALIZED MUSIC - Thought and Mathematics in Music (Revised Edition)*, PENDRAGON PRESS, 1992.
- [45] David Lewin: *Generalized Musical Intervals and Transformations*, Oxford University Press, 2007.
- [46] M. Tsao: *Abstract Musical Intervals: Group Theory for Composition and Analysis*, pp.93-171, Musurgia Universalis Press, 2007.
- [47] Robert Morris: *Composition with Pitch-Classes*, Yale University Press, 1987.
- [48] 一柳慧: 音楽という営み, NTT 出版, 1998.
- [49] J. J. Fux: *Gradus ad Parnassum*, 1725. (坂本良隆訳: 古典対位法, 音楽之友社, 1950.)
- [50] 長谷川良夫: 対位法, 音楽之友社, 1955.
- [51] Noël Gallon, Marcel Bitsch: *Traité de Counterpoint*, Durand & Cie, 1964. (矢代秋雄訳: 対位法, 音楽之友社, 2006.)
- [52] 只野真奈美, 山崎直子, 高田正之, 小谷善行, 西村恕彦: “対位法による編曲システムの開発,” 情報処理学会第 38 回全国大会講演論文集, 4W-2, pp. 1863-1864, 1989.
- [53] 吉川響, 中井満, 下平博, 嵯峨山茂樹: “動的計画法を用いた音楽の対旋律の自動生成,” 平成 12 年電気関係学会北陸支部大会講演論文集, F-82, p. 383, 2000.
- [54] M. Farbood and B. Schonert: “Analysis and synthesis of Palestrina-style counterpoint using Markov chains,” in *Proceedings of the International Computer Music Conference*, pp.471-474, 2001.
- [55] 中瀧昌平, 西本卓也, 嵯峨山茂樹: “動的計画法と音列出現確率を用いた対位法の対旋律の自動生成,” 情報処理学会研究報告 (MUS), 2004-MUS-56, pp. 65-70, 2004.
- [56] 田村理遊, 但馬康宏, 小谷善行: “音高と音価の隠れマルコフモデルを用いた自動副旋律生成,” 情報処理学会研究報告 (MUS), 2007-MUS-69, pp. 7-12, 2007.
- [57] ポール・グリフィス著、石田一志・佐藤みどり共訳: 現代音楽 - 1945 年以後の前衛, 音楽之友社, 1987.

- [58] <http://www.freesound.org/>
- [59] G. Assayag and S. Dubnov: Using Factor Oracles for Machine Improvisation, *Soft Computing*, Vol.8, No.9, pp.604-610, 2004.
- [60] David Cope: *Virtual music: computer synthesis of musical style*, The MIT Press, 2000.
- [61] Michael Edwards: Algorithmic Composition: Computational Thinking in Music, *COMMUNICATIONS OF ACM*, Vol.54, No.7, pp.58-67, 2011.
- [62] C. Agon, M. Andreatta, G. Assayag, S. Schaub: Formal Aspects of Iannis Xenakis "Symbolic Music": A Computer-Aided Exploration of Compositional Processes. *Journal of New Music Research*, 33(2), 2004.
- [63] Y.K. Ahn, C. Agon, M. Andreatta: Structures Ia pour deux pianos by Boulez: towards creative analysis using OpenMusic and Rubato. *Proc. Mathematics and Computation in Music*, Berlin, Germany, 2007.
- [64] Charles de Paiva Santana, Jean Bresson, Moreno Andreatta: MODELING AND SIMULATION: THE SPECTRAL CANON FOR CONLON NANCARROW BY JAMES TENNEY, *Proc. SMC*, pp.431-437, 2013.
- [65] Allen Forte: *The Structure of Atonal Music*, 1973 (森 あかね (訳) 無調音楽の構造, pp.27-44, 音楽之友社, 2012) .
- [66] Richard Cohn, "Neo-Riemannian Operations, Parsimonious Trichords, and Their Tonnetz Representations" , *Journal of Music Theory*, Vol.41/1, 1997.
- [67] 小方厚: 音律と音階の科学, 講談社, 2007.
- [68] William A. Sethares: *Tuning, Timbre, Spectrum, Scale*, Springer-Varlag, London, 2005.
- [69] 日比美和子: 無調音楽の分析理論における「相似性 (similarity)」概念の展開とその意義, *先端芸術音楽創作学会会報*, Vol.1, No.4, pp.1-8, 2010.
- [70] 日比美和子: 無調音楽における声部進行の規則の検証 —シェーンベルク作曲《弦楽四重奏曲第二番 op. 10》より終楽章〈忘我〉へのユージン・ナームアの暗意／実現モデルの適用から—, 東京芸術大学博士論文, 2012.
- [71] 高岡明: ピッチクラス集合論に基づく楽曲分析プログラムの実装と問題点について, *音楽情報科学* 38-3, 2000.
- [72] Clifton Callender: Voice-Leading Parsimony in the Music of Alexander Scriabin, *Journal of Music Theory*, Vol. 42, No. 2, Neo-Riemannian Theory (Autumn, 1998), pp.219-233, 1998.
- [73] <http://essentia.upf.edu/>
- [74] C. Bishop: *Pattern Recognition and Machine Learning*, Springer, 2006.
- [75] D. Cooke: *The Language of Music*, Oxford University Press, 1989.

本論文に関連する発表論文

- (1) Tsubasa Tanaka and Kiyoshi Furukawa: Automatic Melodic Grammar Generation for Polyphonic Music Using a Classifier System, *Proceedings of Sound and Music Computing Conference*, pp.150-156, 2012.
- (2) 田中翼, 古川聖: 多声音楽の旋律スタイルの文法的なモデル化および自動生成, *情報処理学会論文誌*, 54 巻 4 号, pp.1308-1318, 2013.
- (3) 田中翼, 古川聖: 群の生成系としての音程スケール, *先端芸術音楽創作学会会報*, Vol.5, No.1, pp.3-9, 2013.
- (4) Tsubasa Tanaka and Kiyoshi Furukawa: Reinforcement Learning of Musical Scales from Human Emotions, *Proceedings of International Computer Music Conference*, pp.449-452, 2013.
- (5) Tsubasa Tanaka, Hidefumi Ohmura and Kiyoshi Furukawa: Reinforcement Learning Models for Acquiring Emotional Musical Modes, *Proceedings of Sound and Music Computing Conference*, pp.40-47, 2013.
- (6) 田中翼, 大村英史, 古川聖: 感情の表現力をもつ旋法の獲得のための強化学習モデル, *音楽情報科学研究会研究報告*, Vol.2014-MUS-102, No.22, pp.1-6, 2014.

付録

以下に第4章の音程スケールの理論を用いて生成した楽曲「Etude in Interval Scale {5,6,7}」の楽譜を掲載する.

Etude in Interval Scale {5,6,7}

Tsubasa Tanaka
(2013)

$\text{♩} = 94$

mp

7

13

19

25

31

31

Musical score system 1, measures 31-36. Treble clef, bass clef. Key signature: one sharp (F#). Measure numbers 31 and 31 are indicated at the start of the staves.

37

37

mp cresc.

Musical score system 2, measures 37-42. Treble clef, bass clef. Key signature: one sharp (F#). Measure numbers 37 and 37 are indicated at the start of the staves. The dynamic marking *mp cresc.* is present in the bass staff.

43

43

cresc.

Musical score system 3, measures 43-48. Treble clef, bass clef. Key signature: one sharp (F#). Measure numbers 43 and 43 are indicated at the start of the staves. The dynamic marking *cresc.* is present in the bass staff.

49

49

mf

Musical score system 4, measures 49-53. Treble clef, bass clef. Key signature: one sharp (F#). Measure numbers 49 and 49 are indicated at the start of the staves. The dynamic marking *mf* is present in the bass staff.

54

54

Musical score system 5, measures 54-58. Treble clef, bass clef. Key signature: one sharp (F#). Measure numbers 54 and 54 are indicated at the start of the staves.

59

59

Musical score system 6, measures 59-64. Treble clef, bass clef. Key signature: one sharp (F#). Measure numbers 59 and 59 are indicated at the start of the staves.

63

63

mp

Ped. Ped. Ped. Ped. Ped. Ped.

68

68

mf

Ped. Ped. Ped. Ped. Ped. Ped. Ped.

73

73

p *f*

Ped. Ped. Ped. Ped. Ped. *

77

77

p *p cresc.*

Ped. Ped. Ped. Ped. Ped. *

82

82

cresc.

Ped. Ped. Ped. Ped.

86

86

mp

Ped. Ped. Ped. Ped.

90

94

98

102

106

110

114

Musical score for measures 114-117. The system consists of two staves, Treble and Bass. The key signature has two sharps (F# and C#). The time signature is 3/4. The music features a rhythmic pattern of eighth and sixteenth notes with frequent accidentals.

118

Musical score for measures 118-121. The system consists of two staves, Treble and Bass. The key signature has two sharps (F# and C#). The time signature is 3/4. The music features a rhythmic pattern of eighth and sixteenth notes with frequent accidentals. A dynamic marking of *f* (forte) is present in measure 121.

122

Musical score for measures 122-124. The system consists of two staves, Treble and Bass. The key signature has two sharps (F# and C#). The time signature is 3/4. The music features a rhythmic pattern of eighth and sixteenth notes with frequent accidentals.

125

Musical score for measures 125-127. The system consists of two staves, Treble and Bass. The key signature has two sharps (F# and C#). The time signature is 3/4. The music features a rhythmic pattern of eighth and sixteenth notes with frequent accidentals.

128

Musical score for measures 128-130. The system consists of two staves, Treble and Bass. The key signature has two sharps (F# and C#). The time signature is 3/4. The music features a rhythmic pattern of eighth and sixteenth notes with frequent accidentals. A dynamic marking of *ff* (fortissimo) is present in measure 129.

131

Musical score for measures 131-134. The system consists of two staves, Treble and Bass. The key signature has two sharps (F# and C#). The time signature is 3/4. The music features a rhythmic pattern of eighth and sixteenth notes with frequent accidentals.

135

135

fff

138

141

144

fff

smorzando

150

150