

論文 / 著書情報
Article / Book Information

題目(和文)	事例に基づく演奏表情の生成に関する研究
Title(English)	
著者(和文)	鈴木泰山
Author(English)	Taizan Suzuki
出典(和文)	学位:工学博士, 学位授与機関:東京工業大学, 報告番号:甲第4484号, 授与年月日:2000年3月26日, 学位の種別:課程博士, 審査員:
Citation(English)	Degree:Doctor of Engineering, Conferring organization: Tokyo Institute of Technology, Report number:甲第4484号, Conferred date:2000/3/26, Degree Type:Course doctor, Examiner:
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

事例に基づく演奏表情の生成に関する研究

1999年12月

東京工業大学 情報理工学研究科 計算工学専攻

指導教官 徳永 健伸 助教授

97D38089

鈴木 泰山

目次

第1章	序論	1
1.1	背景と目的	1
1.2	本論文の構成	3
第2章	演奏表情生成	5
2.1	演奏表情	5
2.1.1	音楽演奏と演奏表情	5
2.1.2	演奏表情の特徴	5
2.1.3	演奏表情の与え方とその実例	8
2.1.4	楽曲構造と演奏表情	10
2.1.5	楽曲外の要因による演奏表情の変化	12
2.2	関連研究	13
2.2.1	規則に基づく演奏表情生成	13
2.2.2	演奏表情生成規則の学習	15
2.2.3	その他の手法	22
2.3	まとめ	25
第3章	事例に基づく演奏表情生成	27
3.1	事例に基づく演奏表情生成	27
3.1.1	概念	27
3.1.2	基本構成と手順	28
3.2	特徴	34
3.2.1	利点	34
3.2.2	欠点	35
3.2.3	関連研究との相違	36
第4章	要素技術	38
4.1	事例の有効利用	38
4.1.1	事例の不足とその影響	39
4.1.2	旋律断片を単位とする検索	39
4.1.3	検索単位が大きい事例の取り扱い	42
4.1.4	旋律断片の分割方法	43
4.1.5	演奏データの有効利用を考慮した参考事例検索	45
4.2	演奏表情の記述と計算	47
4.2.1	演奏表情の表現	48
4.2.2	演奏表情の大局比率	49
4.2.3	事例に基づく演奏表情生成手法での演奏表情の大局比率	55

4.3	旋律の類似性の評価	57
4.3.1	旋律の類似性の評価	57
4.3.2	全体的な特徴による旋律の類似性評価	58
4.3.3	旋律の特徴量	59
4.3.4	全体的な特徴の類似性の再帰的評価	61
4.3.5	旋律の類似性の評価例	62
4.3.5.1	特徴量	62
4.3.5.2	特徴量の合成	64
4.4	演奏状況	65
4.4.1	演奏状況の表現	65
4.4.2	計算機での演奏状況の記述	66
4.4.3	演奏状況の類似性の評価	67
第5章	演奏表情生成システム	69
5.1	概要	69
5.1.1	Kagurame Phase-I の制約	69
5.1.2	システムが取り扱う情報の種別と表記法	70
5.1.3	Kagurame Phase-I で取り扱う演奏表情	76
5.2	構成	77
5.2.1	入出力と演奏データ	79
5.2.2	演奏表情の生成手順	79
5.2.3	実装	82
5.3	実験	82
5.3.1	準備	82
5.3.2	生成	84
5.4	評価	86
5.4.1	再現能力	86
5.4.2	演奏技術	88
5.4.3	演奏状況による演奏表情の相違	91
5.4.4	計算時間	91
第6章	自動伴奏システムへの応用	94
6.1	自動伴奏システムについて	94
6.1.1	概要と先行研究	94
6.1.2	演奏プラン	97
6.1.3	自動伴奏システムとリハーサル	98
6.1.4	まとめ	100
6.2	事例に基づく演奏表情生成を用いた自動伴奏システム	100
6.2.1	構成	100
6.2.2	システムの動作	102
第7章	結論	106
7.1	まとめ	106
7.2	今後の課題	106

謝辭

109

参考文献

110

目次

2.1	強弱の程度を表す記号（フォルテ，ピアノ，フォルテシモ）	6
2.2	強さ表す記号と強さの変化を表す記号（フォルテシモ，デクレシェンド，ピアノ）	6
2.3	テンポ表示	7
2.4	速度記号（リタルダンド，アテンポ）	7
2.5	音の強さとテンポの変化	9
2.6	拍節構造とグループ構造	11
2.7	タイムスパン還元と延長的還元の例	11
2.8	規則に基づく演奏表情生成の概念	13
2.9	演奏表情生成規則の学習を伴った演奏表情生成	15
2.10	Widmer の構築した音楽認知モデル	16
2.11	白川らの手法での演奏表情生成規則の学習手順	21
2.12	SaxEx の構成	23
2.13	SaxEx の生成手順	24
2.14	SaxEx のデータ構造	25
3.1	事例に基づく演奏表情生成の概念	28
3.2	事例に基づく演奏表情の生成手法の基本構成	29
3.3	演奏データ集の構成	30
3.4	入力情報の構成	30
3.5	類似旋律の検索	31
3.6	重要度の評価	32
3.7	演奏表情の分析	33
3.8	演奏表情の合成	33
3.9	対象曲への適用	34
4.1	演奏対象曲を旋律断片に分割してから検索	40
4.2	旋律の分割による事例の総数の増化	41
4.3	旋律の分割による旋律の類似確率の向上	41
4.4	参考事例から大局的な演奏表情のみを取得	43
4.5	楽曲構成の境界と階層	44
4.6	楽曲構成の例	45
4.7	全体的な特徴のみから参考事例の重要度を評価することの問題	47
4.8	全体的な演奏表情と局所的な演奏表情の分離	47
4.9	演奏表情の変動パターン曲線	49
4.10	演奏表情の大局比率の概念	50
4.11	演奏表情の大局比率で利用する境界情報の条件	50
4.12	演奏表情の大局比率の計算	51

4.13	演奏表情の大局比率を用いた演奏表情の記録	52
4.14	テンポに関する演奏表情の大局比率の計算例	53
4.15	演奏表情の大局比率から絶対的な演奏表情値への復元の例	54
4.16	演奏表情の大局比率の合成	56
4.17	音高の推移の差分を利用した旋律比較	57
4.18	小節や拍などの特徴情報を利用した比較	58
4.19	旋律の類似性の評価手順	59
4.20	旋律の全体的な特徴を利用した再帰的评价	61
4.21	旋律の類似性の評価例（ト音記号，四分の二拍子）	62
5.1	図 5.2 の SMF の楽譜	71
5.2	SMF の例	71
5.3	調，音高，オクターブ値の関係（] の右側の数値がオクターブ値）	73
5.4	楽譜情報と楽曲構成情報の例	75
5.5	演奏状況の記述例	76
5.6	演奏長比	77
5.7	演奏長比と演奏表情の関係	77
5.8	Kagurame Phase-I の全体構成	78
5.9	演奏データと楽譜情報との対応づけ	80
5.10	実験に用いた楽曲構成の一例（8小節の練習曲の 24 番）	83
5.11	演奏データの収集環境	84
5.12	演奏データの録音環境	85
5.13	クローズドテストで生成した演奏と人間の演奏との比較（Czerny の 8 小節の練習曲の 1 番のロマン派風の演奏のテンポ）	88
5.14	オープンテストで生成した演奏と人間の演奏との比較（8小節の練習曲の 37 番のロマン派風の演奏のテンポ）	89
5.15	オープンテストで生成した演奏と人間の演奏との比較（8小節の練習曲の 37 番のロマン派風の演奏の音の強）	89
5.16	オープンテストで生成した演奏と人間の演奏との比較（8小節の練習曲の 37 番のロマン派風の演奏の演奏長比）	90
5.17	オープンテストで生成した演奏と人間の演奏との比較（8小節の練習曲の 1 番のロマン派風の演奏の音の強さ）	91
5.18	オープンテストで生成したロマン派風と古典派風の演奏の比較（8小節の練習曲の 37 番の演奏のテンポ）	92
5.19	演奏状況が中間的な演奏データの比較（8小節の練習曲の 37 番の演奏の音の強さ）	92
6.1	自動伴奏システムの標準的な構成	95
6.2	演奏プランと演奏表情との関係	97
6.3	ツーフレイズリハーサル	99
6.4	事例に基づく演奏表情生成手法を導入した自動伴奏システム	101
6.5	事例に基づく演奏表情生成手法を導入した自動伴奏システムの動作	102
6.6	初めての相手との合奏	104
6.7	演奏経験がある相手との合奏	104
6.8	リハーサル後の合奏	105

表目次

2.1	強弱記号	6
2.2	Widmer のシステムが学習した演奏表情生成規則の内訳	18
4.1	テンポに関する演奏表情の大局比率の計算例	53
4.2	大局比率の復元の結果	55
4.3	評価例の旋律の音階の分布と特徴量ベクトル	64
4.4	「A 氏による明るくはずむような演奏」の演奏状況	66
5.1	SMF の例	71
5.2	調の表記	72
5.3	長調の場合の階名の表記	73
5.4	調ごとの基本音高	74
5.5	階名と音高の対応	74
5.6	実験曲	83
5.7	クローズドテストで用いた演奏状況設定	84
5.8	オープンテストで用いた演奏状況設定	85
5.9	クローズドテストでの人間の演奏とシステムの演奏の差異	87
5.10	システムがクローズドテストで生成した演奏状況の異なる演奏データの差異	87
5.11	演奏表情の生成に要した時間（秒）	93
5.12	演奏表情生成の各段階での所要時間の割合	93
6.1	自動伴奏システムのデータと事例に基づく演奏表情生成手法のデータとの関係	103

第1章 序論

1.1 背景と目的

音楽は、我々が考える以上に、人間の生活に深く根差した重要なものである。おそらく、現在の日本のような社会環境に於いて、一日に一度も音楽を耳にしない人は殆んどいないのではなかろうか。そして音楽が我々に与える影響も非常に大きい。特に音楽が持つ心理的影響は絶大である。音楽演奏を、気分の高揚や、作業の能率の向上、集中力の増進などの目的で利用する人も少なくない。さらに音楽は、映画や小説に比べて圧倒的に短い時間で非常に大きな心理的揺動をもたらすこともある。こういった音楽の心理的な影響力を支えている大きな要因の一つに演奏表情がある。

我々が普段耳にしている音楽演奏は、必ずしも楽譜通りに正確に演奏されているわけではない。実際の人間の演奏ではテンポを微妙に揺らしたり強調したい部分の音を強くするなど様々な演出が行なわれており、これにより聴者は音楽に様々な表情を知覚するのである。楽曲の演奏中にみられるこれらのテンポや音の強さの揺らぎが演奏表情である。テンポの揺らぎや音の強弱以外にも、和音のずれや音の出ている時間の長さ(スタッカート、レガート)、個々の楽器に特有な奏法など様々な要素が演奏表情に含まれる。なお、楽曲の編曲は演奏表情に含まない。編曲とは、楽曲の楽譜上の音符の順序や数を変更することで、楽曲の雰囲気や構成そのものを変更することを指す。それに対して、演奏表情は楽譜上の音符の並びを尊重し、楽譜に対して一切の変更を行なわない。

演奏表情は、音楽演奏を表情豊かで聴者を魅了するようなものに仕上げる上で必要不可欠なものである。演奏表情を極力与えず楽譜通りに演奏した曲は、聞いていても面白みに欠けたものになってしまうことが多く、演奏表情の与えられていない演奏は一般的に良い演奏とはいえないことが多い。また、素人の演奏と高名な演奏家による演奏では同じ曲でも全く異なる印象になることがよくあるが、これは演奏技術そのものの違いもさる事ながら、演奏表情の与え方の違いによるところが非常に大きい。ある意味では、聴衆に深い感銘を与えるのも、逆に退屈を誘うのもこの演奏表情にかかっている。

演奏表情の与え方は楽譜には殆んど記載されておらず、事実上その殆んど全ては演奏者の裁量に委ねられている。演奏表情の与え方は自由度が高いため、極端に非常識なものでない限り許容されてしまう。そこで、演奏者はその知識や経験、趣向に基づいて、その演奏曲にはどういった演奏表情が適しているのか、また、いかにして演奏を行なえば聴者により深い感銘をもたらすことが出来るのかを探求し、その結果、各々が独自の演奏表情を与えることとなる。この過程に於ける人間の思考活動は、直観や経験によるところが多く、またその多くは無意識に行なわれているので、演奏表情の生成過程や根拠の大部分は演奏者自身にとっても不明確なものである。さらには、演奏者の心理状態やその場の雰囲気などの楽曲外の要因が演奏表情に与える影響も存在する。なお、本論文ではこれら演奏表情に影響を与える楽曲外要因を演奏状況と呼ぶ。このように、演奏者は多くの知識から複雑な過程を経て演奏表情の付与を行なっている。

近年、コンピュータなどの電子機器の普及にともなって、コンピュータを利用した音楽演奏が広く普及するようになった。通信カラオケに代表される電子音楽再生機器やコンピュータゲームで奏でられている背景音楽がこの好例であろう。コンピュータによる音楽演奏は、従来からある CD や MD などのレコーディングメディアによる演奏とは大きく異なるものである。レコーディングメディアが、事前に録音された音データを直接再生するのに対して、コンピュータによる音楽演奏では楽譜を基に作成された演奏データを、演奏時にリアルタイムに音情報に変換し再生する。従って、コンピュータによる音楽演奏では、再

生時に音量やテンポを変化させたり、音色を変更するといったことが可能である。これは、通信カラオケのキー修正機能（演奏時の調を変更することで音高を変える）やテンポ修正機能に 응용されている。

コンピュータによる音楽演奏では、演奏データの作成に音楽演奏の技術が必要ないため、音楽演奏の経験がない人でも容易に演奏データを作成し、好みの楽曲の演奏を楽しむことができる。また、基本的に楽譜情報しか必要としないので、レコーディングメディアに比べて演奏データの情報量が圧倒的に少なくてすむ。このように、コンピュータによる音楽演奏には、従来の形態の音楽媒体にはない多くの利点がある。

しかしながら、その一方で現状のコンピュータによる自動演奏の評価は、無表情で面白みにかけていて人間の演奏には遠く及ばないというものが一般的である。こういった評価は、コンピュータによる音楽演奏の大部分が演奏表情を伴わないことに起因する。コンピュータによる音楽演奏では、演奏データとして楽譜情報を利用することが多い。先にも述べたように、楽譜には演奏表情の与え方が殆んど記載されていないため、楽譜情報に基づいた演奏データを直接再生しただけでは演奏表情は得られない。しかしながら、演奏表情を付与するためには音楽経験や直観といった高度な音楽知識が必要になるため、音楽に対する知識や直観を持たないコンピュータが自ら演奏表情を生成し演奏することは困難である。

コンピュータによる音楽演奏において演奏表情を伴った演奏を行なうためには、演奏データの作成者が楽譜情報を入力する際に演奏表情を同時に入力する方法が有効である。ところが、演奏表情を入力するためには、演奏データの作成者に高度な音楽知識が要求される。また、演奏表情の中には無意識に行なっている部分も多く、直観的な曖昧なものを数値化する作業になるため、演奏表情を伴った演奏データを作成するのは非常に根気のいる作業である。このような理由から実際には演奏表情を伴った演奏データは殆んど流通していないのが実状である。

コンピュータによる音楽演奏に演奏表情を持たせるもうひとつの、そして本質的な解決策は、コンピュータ自身に演奏表情生成能力を与えるというものである。既に述べたように、演奏表情の生成には音楽経験や直観といった音楽知識が必要不可欠であり、この知識の複雑さがコンピュータによる演奏表情生成の大きな妨げとなっている。しかしながら、このような複雑さ故に人工知能や認知科学の観点から見ると、演奏表情の自動生成は非常に興味深い課題となりえる。こうした状況を反映して、コンピュータ上での音楽知識のモデル化や、それを利用した演奏表情の自動生成の研究も数多く行なわれるようになってきた。

もともと、音楽演奏教育や音楽心理学の分野では古くから演奏表情の生成の規則やメカニズムの分析が行なわれていた。そして、旋律の特徴とそこに見られる典型的な演奏表情との間に様々な法則や傾向があることが知られていた。演奏表情の生成研究はこれらの法則を利用することから始まった。演奏表情生成規則という形でコンピュータ上に実装されたこれらの音楽演奏知識を、演奏対象曲に適用して演奏表情を生成するというものである。さらには、最近では、人間の演奏から演奏表情生成規則を自動獲得するという研究も数多く見受けられる。

先にも触れたが、演奏表情の与え方は同じ楽曲でも演奏者によって全く異なる上に、たとえ演奏者が同じでも演奏者自身の心理状態や演奏者を取り巻く環境の影響によって様々な変化する可能性も持っている。また、聴取する側の演奏表情の好みも人によってまちまちである。従って、演奏表情生成手法にも、演奏者や演奏時の環境などの演奏状況を考慮し、同一の楽曲に対して多種多様な演奏表情を生成するような能力が望まれる。

しかしながら、先行研究ではこの演奏表情の多様性はあまり考慮されていない。事実、これらの演奏表情生成手法は単一の楽曲に対して一種類の演奏表情しか生成できないものが殆んどである。

本研究では、多様な演奏表情の生成が可能な手法として、事例に基づく演奏表情生成手法の提案、実装、評価を行なう。本研究の目的を以下に示す。

- 事例に基づく演奏表情生成手法を提案する。
- 事例に基づく演奏表情生成手法を実現する上で必要となる要素技術について提案する。これは、演奏データ集の有効利用技術、演奏表情の分析・記述法、旋律の類似性の評価方法、演奏状況の記述と類

似性の評価に関するものである。

- 事例に基づく演奏表情生成手法を演奏システムとして実装し、事例に基づく演奏表情生成手法の有効性や、演奏状況に応じた演奏表情の生成能力について検証する。

事例に基づく演奏表情生成手法は、演奏者や演奏環境など演奏表情に影響を与えるような楽曲外の要因を考慮し、単一の楽曲に対してこれらの要因に応じた多様な演奏表情の生成を可能にする演奏表情性正手法である。

先行研究で演奏状況が十分に考慮されず、単一の演奏表情しか生成できないものが多いのは、先行研究が共通して用いている手法に原因がある。先行研究の多くは、演奏表情の生成に演奏表情生成規則を利用している。規則を用いる手法で多様な演奏表情の生成を行なうためには、演奏表情の生成規則に楽曲と演奏表情との関係だけではなく演奏状況と演奏表情の関係も導入する必要がある。しかし、演奏状況は数多くのパラメータの組合せで構成されており、また個々のパラメータと演奏表情との関係も楽曲の特徴と演奏表情との関係ほどは明確ではない。元来、規則というアプローチは多様な要因と連続的に変化するパラメータに馴染みにくく、これが先行研究に演奏状況を導入する妨げとなり、結果として多様な演奏表情の生成ができなかったと考えられる。

本研究で提案する事例に基づく演奏表情は、音楽演奏知識として従来の演奏表情生成規則の代わりに、人間による音楽演奏の演奏データ集を利用する。演奏対象曲の楽譜情報を入力すると、演奏データ集の中から演奏対象曲に類似した楽曲の演奏データを検索し、その演奏データの中で人間がどのような演奏を行っていたかを分析して、それをそのまま演奏対象曲に適用する。演奏対象曲に適用する際には、対象曲と演奏データとの類似性を評価し、より似ている曲に見られる演奏表情を重視する。この類似性の評価には、単に楽曲の類似性だけではなく、演奏データの演奏状況の類似性を考慮することも可能である。そこで、演奏データごとにあらかじめ演奏時の演奏状況を付記しておき、また、入力として生成したい演奏表情の演奏状況設定を与える。そして、演奏データの類似性の評価の際には、入力された演奏状況設定に近い演奏状況の演奏データをより重視する。これによって、同じ楽曲に対して演奏状況に応じた多種多様な演奏表情が生成できる。

事例に基づく演奏表情生成手法を実現するためには、様々な要素技術が必要となる。事例に基づく手法特有の問題として事例の不足がある。これは、演奏対象曲に類似した演奏データの検索がうまく行かず、参考になる演奏データが殆んど得られないと、演奏表情生成が全く行なえないというものである。この問題は、本手法の根幹に関わるものであることから、可能な限り事例の不足の発生を抑制するような技術が必要となる。そこで、事例の不足を抑制する技術として、演奏データ集の有効利用法について提案する。この技術を導入するためには、演奏表情の記述や表現手法や、演奏データの検索技術にも従来とは違ったアプローチが必要となるため、これらについても検討、提案を行なう。また、従来の研究では演奏状況についてあまり考慮されていなかったため、演奏状況の表現やその類似性の評価について検討する。

さらに、事例に基づく演奏表情生成手法の有効性を検証するためには、実際にこの手法を用いた演奏システムを計算機上に実装する必要がある。そこで、この手法を用いた演奏表情生成システム Kagurame を構築する。また、このシステムを利用した演奏データの生成実験を行ない、その演奏データの評価を通して、事例に基づく演奏表情生成手法の演奏表情生成能力や演奏状況への対応能力について検証、考察する。

1.2 本論文の構成

本論文の構成は以下の通りである。

第2章では、演奏表情およびその生成について説明する。前半では、音楽演奏に於ける演奏表情の役割や、人間がいかにして音楽演奏に演奏表情を付与しているかについて述べる。第2章の後半では、演奏表情の分析や生成に関する先行研究を紹介すると共に、その多くが抱える問題点を分析する。

第 3 章では、事例に基づく演奏表情生成を提案する。まず、事例に基づく演奏表情の生成手法の概念や構成、手順について詳説する。また、この手法の特徴や先行研究との比較について概観する。

第 4 章では、事例に基づく演奏表情の生成を実現するにあたって必要となる要素技術を提案する。第 4 章の各節では、1) 演奏データ集の有効利用に関する技術、2) 演奏表情の記述、分析、計算法、3) 旋律の類似性の評価手法、4) 演奏状況の記述、の 4 つの要素技術について詳説する。

第 5 章と第 6 章の 2 章では事例に基づく演奏表情生成手法の応用について報告する。第 5 章では、事例に基づく演奏表情生成手法を利用した演奏データ生成システム Kagurame について述べる。また Kagurame を用いた演奏表情の生成実験を報告すると共に、この結果から事例に基づく演奏表情の生成手法についても評価を行なう。第 6 章では、事例に基づく演奏表情生成手法を応用した自動伴奏システムを紹介する。ここでは、このシステムの特徴、構成、実装について述べる。

最後に第 7 章で本論文のまとめと今後の課題について述べる。

第2章 演奏表情生成

2.1 演奏表情

2.1.1 音楽演奏と演奏表情

人間による音楽の音楽演奏は必ずしも楽譜通りに正確に行なわれているものではない。勿論、音符の順序や和音の組合せなど大部分は楽譜に従って演奏するが、一方で、音の強さや演奏のテンポは必ずしも楽譜通りではなく、殆んどの場合は楽譜に規定されている値から逸脱したり様々に変化したりする。

音の強さやテンポの変化は、音楽演奏に様々な雰囲気を与えるという効果がある。例えば、テンポが速く音の強い演奏は攻撃的で躍動感のある演奏になるし、逆にテンポが遅く音も弱い演奏は沈んだ雰囲気を演出する。音の強さやテンポなど、音楽演奏中に見られ、演奏の雰囲気を形成するのに必要な要素を演奏表情と呼ぶ。楽譜からの逸脱が観測されるのは、この演奏表情と呼ばれる要素である。

演奏表情の与え方は楽譜上に明確に規定されているわけではなく、殆どが演奏者の裁量に委ねられている。演奏者は、演奏曲によって表現しようとしている内容に応じて、音の強さやテンポを様々に変化させ、聴衆に間接的に雰囲気を伝え深い感銘を与えるのである。しかしながら、楽曲に対する解釈やその表現方法は演奏者に応じて多岐に変化するため、演奏表情の与え方も演奏者に応じて様々に変化する。同じ楽曲の演奏であるにも関わらず、演奏から受ける印象が演奏者によって大きく変わるのは、演奏技術もさることながら、演奏者によって演奏表情の与え方やその技量の相違によるところが大きい。

2.1.2 演奏表情の特徴

本項では演奏表情について概説する。単に演奏表情と言っても様々な要素がある。代表的なものを以下に示す。

- 音の強さ
- テンポ
- 演奏長
- 和音のずれ
- 楽器ごとの奏法

以下でこれらの演奏表情について概説する。

音の強さ 音の強さは演奏表情の中では単純なものである。ピアノや木琴、鉄琴、打楽器など、音源に衝撃を与えることによって生じる振動を音として利用するような楽器の場合は、衝撃の強さに応じて音の強さが変化する。弦楽器などの摩擦による共振現象を利用する楽器の場合は、摩擦のエネルギーによって共振の振幅をコントロールし、音の強さを変化させることができる。管楽器やオルガンなどの空気による共鳴を利用した楽器の場合は、空気の圧力や流速に応じて音の強さが変化する。衝撃を利用する楽器の場合

は、瞬間的な衝撃によって発音するため、一旦衝撃を与えると、振動が減衰するに従って音の強さは弱くなっていく。共振や共鳴を利用している楽器の場合は摩擦や振動を連続的に入力する必要があるので、発音中に入力の強さを変化させることにより動的に音の強さを変化させることが可能となる。なお、発音後の音の強さの変化は、演奏表情の分類上、音の強さではなく後述する楽器ごとの奏法に分類する。

音の強さの与え方は楽譜上にその指針が表示されていることが多い。楽譜中には、フォルテ、ピアノなどの大まかな強さを表現する記号（図 2.1）や、クレッシェンド、デクレッシェンド（ディミニエンド）のように連続的な変化を表す記号（図 2.2）が頻繁に見られる。これらの記号を強弱記号と呼ぶ [36]。主要な記号を表 2.1 に示す。



図 2.1: 強弱の程度を表す記号（フォルテ、ピアノ、フォルテシモ）



図 2.2: 強さ表す記号と強さの変化を表す記号（フォルテシモ、デクレッシェンド、ピアノ）

強さの程度を表す記号	<i>ff</i> （フォルテシモ）、 <i>f</i> （フォルテ）、 <i>mf</i> （メゾフォルテ）、 <i>mp</i> （メゾピアノ）、 <i>p</i> （ピアノ）、 <i>pp</i> （ピアニシモ）、など
強さの変化を表す記号	<（クレッシェンド）、>（デクレッシェンドもしくはディミニエンド）

表 2.1: 強弱記号

強弱記号は音の強さを絶対的に規定するものではない。フォルテやピアノなどの程度記号は音の強さを相対的に指示しているだけである。さらに、それぞれの記号の間の音の強さの比率も規定されていない。よって、強弱記号から絶対的な音の強さを決定することはできない。

また、同じ強弱記号の部分がすべて同じ強さというわけでもない。例えば、フォルテが表示されている区間でも場所によって音の強さが変化することがあるし、同じフォルテが表示されている他の場所とも違う強さになることが多い。フォルテが表示されている区間の音の強さが、他のフォルテシモ（フォルテより強く演奏するという意味の記号）が表示されている部分より強くなるような逆転現象も発生する。また強弱記号の境界で強さをどう変化させるかも規定されていない。だんだん強さを变化させた方が良い場合もあるし、逆に急激に強さを变化させた方が良い場合もある。クレッシェンドなどの変化記号も同様で、変化曲線の形状や変化量は演奏者の判断に委ねられている。

楽譜上の強弱記号は、あくまで演奏の指針として与えられているにすぎず、必ずしもその指示に従わなければならないというわけではない。実際に音の強さをどのように変化させるかの判断はすべて演奏者に任されている。演奏者によっては、実際の演奏音の強弱と楽譜上の強弱記号とが逆になることも珍しくない。

テンポ テンポは演奏の速度のことである。楽譜上の音符は、音符同士の長さの比率を規定しているだけで、実際にそれぞれの音符を何秒間演奏するかは規定していない。例えば、四分音符は八分音符の二倍の長さで演奏するのが標準であるが、四分音符を何秒間演奏するかは楽曲によって異なる。従って、音符の演奏長から実際の演奏時間を決定するためには、基準となる音符の演奏長を何秒とするかを表す指標が必要である。この指標がテンポである。

テンポは楽譜の左上に記されていることが多い。通常は、四分音符や八分音符など基準となる音符の一分間の演奏回数で表記する。図 2.3 の例では、四分音符 = 180 と表記されているが、これは一分間に四分音符を 180 回演奏する速度という意味である。従って、この例の場合、四分音符一個の演奏時間は $60/180 = 1/3$ 秒となる。テンポの数値が大きくなれば演奏が速くなり、逆に小さくなれば遅くなる。



図 2.3: テンポ表示

楽譜によっては、具体的な数値ではなく *poco Adagio* (すこし遅く) や *molto Allegro* (非常に速く), *Allegretto* (やや速く) の様に、イタリア語の形容詞や副詞でテンポを表示していることもある。

曲中では、リタルダンド (*rit.*) やアツツエレランド (*accel.*) といったテンポの変化を表す指標や、元のテンポに戻るという意味のアテンポ (*a tempo*), 長く伸ばすという意味のフェルマータなどを用いて、テンポの変化を表すことがある (図 2.4)。いずれも音の強さと同様に、速度の相対的な変化を表しているだけで、変化曲線や変化量は演奏者に全てまかされている。また、これらの指標は必ず守らなければならないというものでもない。



図 2.4: 速度記号 (リタルダンド, アテンポ)

演奏長比 演奏長比は実際に楽譜上に記されている音符長に対して、どれだけの長さ発音したかのことである。実際の演奏では、楽譜上に書かれた音符の長さを正確に発音しているわけではない。音符の最後の部分で若干早めに発音を停止し、次の音との間に休みをおくことが多い。

これらの中には、鍵盤楽器の運指の都合や管楽器の息継ぎなど、技術的に音を継続することができないために生じる休みもあるが、演奏表情として意図的に行なわれる。この例として、スタッカートやテヌート、レガートがある。スタッカートは楽譜上の音長に比べて極端に短く区切って演奏することである。スタッカートでは、音符の長さの大部分の時間が無音時間になる。はずむようなリズムカルな雰囲気を与えることができる。テヌートはスタッカートの逆で音符の長さを十分に保って演奏することを表す。また、レガートは楽譜上の音長より長く発音し微妙に次の音と重ねる演奏法である。流れるような優雅さを表現する時などに使われる。

スタッカートやレガートのような顕著な演奏長の変化は楽譜上に記号として示されているが、それ以外の部分についても曲想などに応じて様々に変化させて演奏することがある。

和音のずれ 楽譜上で同じ拍到複数の音符が並んでいるものを和音と呼ぶ。楽譜上では演奏の時刻は完全に同じ拍にあるが、実際の演奏では同時に演奏を行なうとは限らない。

これも演奏長と同様に運指の都合や物理的な制約により必然的にそうになってしまう場合もあるが、和音の各音を意図的に若干ずらして発音することもある。例えば、低い音から高い音に向かって順番に発音する奏法や、高い方から低い方に順番に発音する奏法がある。これらは、華やかさを演出する時などに使われる。楽譜上に記号で示されている場合が多いが、楽譜には何も記載されていなくても若干ずらし気味に演奏することもある。

楽器ごとの奏法 その他、楽器ごとに固有の演奏技術にも演奏表情に分類するものがある。音の強さでも簡単にふれたが、弦楽器や管楽器などの持続音楽器における発音中の音の強さの変化がその好例である。他にも代表的なものとしてビブラートがある。ビブラートは主に弦楽器で用いられる奏法で、演奏音の高さを微妙に変化させるというものである。通常は、半音程度の音域の間で周期的に音高を振動させる。一連の発音の中で振動の周期や振幅を変化させることもある。ビブラートを与える場所や振幅、周期は演奏者が各々で決める。

2.1.3 演奏表情の与え方とその実例

前項で概説したように、演奏表情には様々な要素がある。これらの要素はそれぞれが別個に単独で与えられることもあるが、通常は複数の演奏表情を組み合わせて与えられることも多い。例えば、テンポ徐々に遅くしながら同時に音を弱くして終末感を演出したり、演奏長比を短くしつつ強く演奏することで個々の音符を際立たせるのが好例であろう。また、演奏表情は特定の箇所だけに与えられるわけではなく、楽曲の最初から最後まで至るところで見られる。顕著な演奏表情が見られないように感じる場所でも、実際には非常に細かな演奏表情が与えられていることが多い。

図 2.5 に実際の人間の演奏中に見られる音の強さとテンポの変化の一例を示す。これは 8 小節の練習曲に対して人間が演奏を行なったものである。演奏自体は典型的なルバートのかかった演奏である。このグラフは、楽曲の一音ごとに、テンポと音の強さの平均値に対する比率を表したものである。横軸は小節、縦軸は平均に対する比率を表す。なお、この曲は 5.3 節で述べる実験で使用した演奏データの一つである。

このグラフからも、テンポや音の強さが楽曲の全体に渡って常に何らかの変化をしていることがわかる。その変化の幅が大きく、例えば、テンポは平均値の半分程度（4 小節目後半）から 2 倍（楽曲の最後）まで、最大で 4 倍以上もの変化をしている。また、前半部分ではテンポと音の強さが連動して変化しているのに対し、後半部分ではテンポの変化と音の強さがちょうど対をなす様に変化しているのが特徴である。

演奏表情の与え方には、数多くの標準的なパターンがあることが知られている。その中でも代表的なものを以下に示す。

- 楽曲やフレーズの終末部分のテンポは徐々に遅くなる
- 音高が徐々に上がっていく部分では、音の強さもそれに従って強くなる
- 小節の一拍目は強めに演奏する
- 周囲に比べて音高が高い音は長めに演奏する

これらの経験則は初心者には演奏法を説明する際などによく用いられる。しかし、これらのパターンを守らなければならないというものではない。これらのパターンを守っていればそれらしい演奏に聴こえるこ

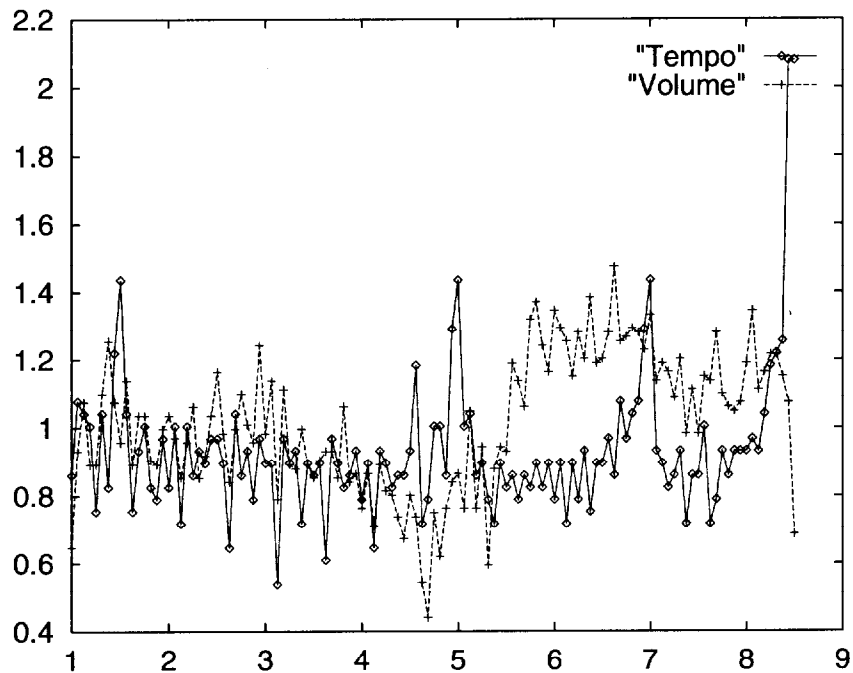


図 2.5: 音の強さとテンポの変化

とが多いが、楽曲によってはこのパターン通りの演奏が適切ではないこともある。経験を積んだ演奏家は、これらのパターンにとらわれず、自分の感性に従って自由に演奏を行なっている。その結果として、これらのパターンに合致する演奏になることもあるが、最初からこれらのパターンを意識して演奏表情を決定しているわけではない。また、いずれも定量的な表現ではないため、強さやテンポをどの程度変化させるかや、どこからどこまでに適用するかは、これらの表現から推し量ることはできない。実際には、具体的な数値は楽曲の特徴に応じて様々に変化する。

上記のパターンの例もそうであるが、演奏表情は単一の要因にしたがっているわけではなく複数の要因が合わさっており、それぞれの要因の影響範囲や依存範囲は、局所的なものから大局的なものまで様々である。上記のパターンを例にとると、「一拍目の音を強くする」は影響範囲は一拍目の音符のみで、また1小節分の旋律にしか依存しない。しかし、「周囲に比べて音高が高い音は長めに演奏する」の場合、影響範囲は音高が高い音の一音だけであるが、周囲のいくつかの音符の音高に依存しており、その範囲の長さが一定ではない。「音高が徐々に上がっていくと音の強さも強くなる」の場合、音高が上がっている区間の全ての音符がお互いに依存して影響を与えている。この場合も範囲の長さは一定ではない。このように、実際の演奏表情は音符がお互いに依存し、全体の流れや部分部分での変化が複雑に絡み合って成立している。

演奏の雰囲気は、演奏に与えられた演奏表情の数値や範囲の微妙な変化によって大きく変わる。経験豊富な演奏者は、演奏表情の微妙な変化による雰囲気の違いを利用して、楽曲が表現しようとしている内容を表現する雰囲気を、演奏の中で意図的に作り出すことができる。このような演奏表情が持っている雰囲気は、しばしば形容詞や比喩表現を用いて言葉で表現される。この演奏表情の雰囲気を表す表現を曲想と呼ぶ。例えば、「明るく」、「はずむように」、「行進曲風に」が曲想表現である。それぞれの曲想表現に対し、どのような演奏表情の与え方を行なうのが適切かについても様々な経験則が知られている。例えば、「はずむような演奏」では、音の強さを全体的に強めにし、テンポの揺らぎを少なく、一音ごとの演奏長比は短めにスタッカートで演奏する。こういった曲想表現が楽譜上に記されていることもあり、演奏表情を決定するための一つの要因となっている。

2.1.4 楽曲構造と演奏表情

演奏表情と楽曲の認知構造には密接な関係があり、類似した認知構造を持つ旋律の演奏表情は類似する傾向があることが指摘されている。認知的楽曲構造とは、人間の音楽の認知過程に共通して見られる傾向を分析し理論化したものである [66]。認知的楽曲構造としては Lerdahl と Jackendoff による調性音楽の生成理論 [42] や Narmour の暗意と実現のモデル [46] が有名である。本項では、Lerdahl らの調性音楽の生成理論について概説する。

楽曲構造 Lerdahl らの調性音楽の生成理論は、人間の音楽認知過程に基づいて音楽中の階層的な構造を解析する手法である。この解析手法では次に挙げる 4 つの構成要素に注目して楽曲構造の解析を行なう。

- グループ構造

グループ構造解析では、楽曲を音楽的なまとまり感があるグループ分割し、その間の階層構造を決定する。グループ構造の解析にはグループ構成規則とグループ選好規則の 2 種の規則を利用する。グループ構成規則はグループ構造が満たさなければならない制約である。例えば、グループの中に含まれる複数のサブグループが交差してはいけない、といったものがある。グループ選好規則はグループの境界の剪定方法に関するヒューリスティクスである。

- 拍節構造

拍節構造は楽曲中の拍の相対的な強弱の関係を表す。拍節構造もグループ構造と同様に階層性がある。拍節構造の解析には拍節構成規則と拍節選好規則の 2 種の規則を利用する。それぞれの規則の役割はグループ構造と同様である。グループ構造と拍節構造の例を図 2.6 に示す。この図では“(”で括られている部分がグループ構造，“[”で括られている部分が拍節構造をあらわす。また，“.”がついている拍は拍節内の強拍である。

- タイムスパン還元

タイムスパン還元と次の延長的還元の二つは還元理論と呼ばれている。これらの解析は楽曲の中から互いに関連している部分を見つけ出し、その依存関係や抽象的な役割を決定するものである。楽曲を構成する音符にはそれぞれ重要度や主従関係といったものがある。例えば、装飾音は後に続く被装飾音符に従属するものであり、楽曲に対して装飾音が持っている役割はそれほど大きくない。タイムスパン還元の木構造はこのような楽譜中の音符間の主従関係を階層的に表現する。図 2.7 にタイムスパン還元の木構造の一例を示す。木構造の枝の部分は、幹の部分に従属しており、幹の方が枝より構造的に重要であることを表している。例えば、第 1 小節の一拍目の音符にはその次の 16 分音符と 8 分音符が従属している。また、16 分音符は後の 8 分音符に従属している。タイムスパン還元も構成規則と選好規則の 2 種の規則から決定する。これらの規則はグループ構造や拍節構造、音符情報を利用して木構造をボトムアップに抽出する。

- 延長的還元

延長的還元は緊張と弛緩の認知構造とも呼ばれる。Lerdahl らは、楽曲は中心となる部分に向かって徐々に緊張度を高めていき、そこを過ぎると今度は終末部に向かって緊張感を解決し弛緩していくという抽象的な構造を持っていると主張している。楽曲中では細かな緊張と弛緩を絶えず繰り返しており、この緊張/弛緩構造が楽曲のまとまり感の根源になっている。延長的還元はこのような楽曲中の緊張と弛緩の構造の抽出に関する理論である。緊張/弛緩構造も木構造で表される。図 2.7 は延長的還元構造の情報も含んでいる。木の幹が左下がりになっている部分が緊張、右下がりの部分が弛緩を表す。図 2.7 の場合、第 1 小節と第 2 小節が緊張部分で、第 3 小節と第 4 小節が弛緩部分である。

また、木の最上部が左下がりになっていることから、この旋律はこの後に続く譜例には記されていない旋律に対して緊張部分となっていることがわかる。



図 2.6: 拍節構造とグループ構造

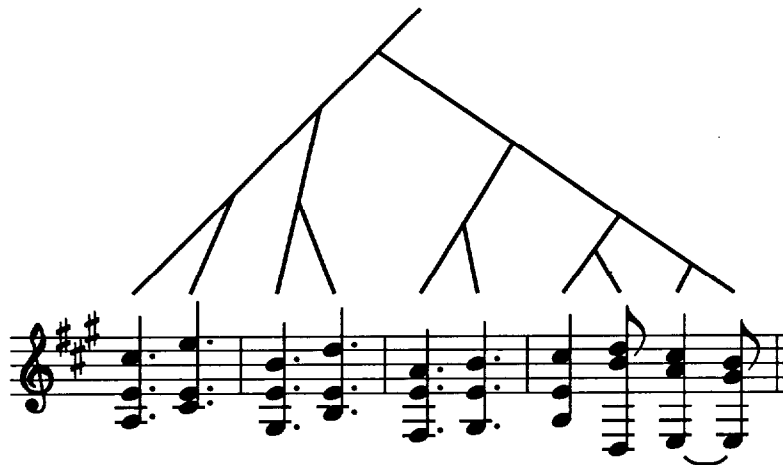


図 2.7: タイムスパン還元と延長的還元の例

グループ構造と拍節構造の理論は調性音楽の生成理論以前にも同様のものが数多く存在した。Lerdahl らの理論はそれらをまとめて体系化したものである。タイムスパン還元と延長的還元は楽曲の表層的な構造から抽象的ないわば意味に相当する情報を取り出し、楽曲の他の部分との関連づけを行なうものである。これらは、調性音楽の生成理論に固有のものである。

認知的楽曲構造に関する理論は音楽学の専門家によるものであるため、その解析手法はアルゴリズムックではない。そこで、理論では曖昧な部分を明確にし、計算機上に解析手法を実装する研究も行なわれている [38, 43, 70, 78].

楽曲構造と演奏表情との関係 楽曲構造と演奏表情とは密接な関係があるといわれている [41, 65]. その論拠は様々であるが、「演奏表情は楽曲の構造を表現し、聴衆に伝えるために付与されているから」という主張が的を射た理由ではないかと思われる。これは、我々が日常の会話に於いて言葉に抑揚を与えることで文章の構造を表現しようとしているのと同様に、音楽演奏者も音楽の演奏に演奏表情という抑揚を与える

こと楽曲の構造を表現しようとしているということである [20]。上記の楽曲構造解析が示すように音楽は内在的な構造を持っているが、これらの主張では、おそらくは、楽曲の構造には様々な曖昧性があるため演奏者が演奏表情を用いて構造を補完し表現することで、聴衆により確実な安定感をもたらすことが演奏表情の一つの役割だと捉えてるのではないかと思われる。自身の演奏の聴取の経験に照らし合わせてみても、演奏表情には少なからず楽曲のまとまり感を演出する役割がある点は否めない。この主張が正しいと仮定すれば、演奏表情が楽曲の構造と密接な関係を持っているのも至極当然のことと言えよう。

こう考えてみると、先に述べた調性音楽の生成理論も実際の演奏における演奏表情の変化と符合する点が多い。例えば、拍節構造はそれ自体が音符の強弱を決定する構造である。グループ構造の境界では、演奏中でテンポを緩めたり、音の強弱を際立たせるなどの傾向が見られる。これはグループ構造と結び付けて考えるよりも、グループ内の緊張/弛緩構造を表現していると考えの方が自然かもしれない。緊張部分では徐々にテンポを速め、弛緩部分では逆に緩めるといった傾向が見られることから、グループ構造内の緊張/弛緩構造に従ってテンポを変化させると結果的にグループ構造の境界部分でのテンポが際だって緩やかになるとも考えられる。

2.1.5 楽曲外の要因による演奏表情の変化

演奏表情は、楽曲のみから一意に決定できるものでなく、楽曲とは直接的には関係ない様々な外的要因の影響を強く受ける。例えば、演奏者がその好例である。同じ楽曲を演奏しても、演奏者によって演奏表情の与え方が様々であることは既に述べた通りであるが、演奏者は楽曲と直接関係ある情報ではないため、これは演奏表情に影響を与える楽曲外の要因である。これらの楽曲外の要因を総合して演奏状況と呼ぶ。楽曲外の要因は様々なものがある。これらを大別すると、以下に挙げる四種に分類できる。

- 演奏者に関する要因

演奏表情は演奏者の判断で与えられることから、演奏者の趣向や直観に強く依存することから、演奏者が誰であるかということが演奏表情の重要な要因の一つである。演奏者の演奏表情の与え方は、演奏者の演奏経験や師事した演奏者など演奏者の経歴によって形成されることから、これらの要因も演奏者に関する要因の一部と考えられる。

- 曲想に関する要因

2.1.3 項で述べたように、演奏表情は曲想によって分類、表現されることが多い。楽譜で指示されている曲想がどういったものであったかや、演奏者が演奏時にどのような曲想を意図したかも演奏表情を決定する要因の一つである。曲想に関する要因の中には演奏スタイルも含まれる。

- 楽曲に付随する情報に関する要因

楽曲に付随する情報とは、演奏曲の楽譜上に記載されている情報ではなく、演奏曲に関する背景知識のことである。例えば、演奏曲の作曲者や作曲時期、楽曲の様式 [41]、作曲にまつわる逸話などがそれである。また、演奏曲に歌詞がある場合は、歌詞の内容も演奏表情に関係する。

- 演奏時の状況に関する要因

同じ演奏者の演奏であれば、必ず演奏表情が同じになるかというとは必ずしもそうではない。演奏表情は演奏者の演奏時の心理状態の影響を強く受けるため、毎回違った演奏表情になる。従って、演奏者の心理状態に影響を与える要因も演奏状況の一つである。演奏者の心理状態に影響を与える要因としては、演奏を行なった場所や使用した楽器、前後に演奏した楽曲などの演奏に直接関係した事象や、天候、季節、時間などの直接的には関係がない事象がある。

ここに例示しただけでもかなりの数の要因があるが、実際にはこの他にも様々な要因が考えられる。また、要因の中には演奏者の経験や演奏場所のように属性値が不特定多数のものや、曲想表現のように連続的に変化するものがある。演奏表情には、これらの演奏状況が複合的に作用するため、演奏状況と演奏表情との関係を分析することは困難である。

2.2 関連研究

演奏表情生成とは、与えられた楽曲（楽譜）に対して演奏表情を付与し、自然な音楽演奏を生成することである。計算機を利用した演奏表情の自動生成の試みは古くから行なわれ、様々な手法が提案されている [22]。

演奏表情生成研究の究極的な姿は、人間と同じような感性や表現能力を持った音楽演奏の知識モデルの構築にある。といっても、研究ごとにその目的は様々であるが、大別すると工学的動機による研究と認知科学的動機による研究の二種がある。工学的な動機による研究の多くは、最終的な目標を演奏表情の生成を行なうことが可能な自動演奏システムの構築に置いている。こうしたシステムを電子ピアノやコンピュータミュージック機器に導入すると、いわば仮想的なミュージシャンを実現することができる。認知科学的な動機による研究には、人間が音楽を演奏する際の音楽認知行動のメカニズムの解明を目標とした研究が多い。様々な音楽認知理論に基づいて、人間の音楽認知モデルを計算機上に実装し、実際にそのモデルが生成した演奏を評価することによって、音楽認知モデルの評価を行なう研究や、学習機構を利用して楽曲の特徴情報と演奏表情との関連づけを行なうことによって、人間の演奏知識モデルを獲得しようとする試みがあげられる。

本節では、過去に提案された演奏表情の生成手法について紹介し、その傾向や特徴を分析する。

2.2.1 規則に基づく演奏表情生成

古典的に用いられていた手法は、演奏表情の生成規則を利用するというものである。音楽演奏教育や音楽心理学の分野では古くから演奏表情の生成の規則やメカニズムの分析が行なわれていた。1970年代に入ると、計算機を利用した音楽演奏技術が確立され、それにともなって計算機を利用した演奏表情生成の研究が始まる。初期の手法は、古くから知られていた演奏表情の規則や特徴を計算機上に実装したものが多く、

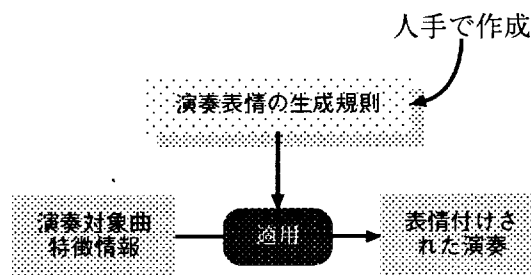


図 2.8: 規則に基づく演奏表情生成の概念

規則に基づく手法の基本的な構成を 図 2.8 に示す。規則に基づく演奏表情生成とは、演奏表情に関する知識を演奏表情生成規則という形で付与し、この規則を用いて演奏表情の生成を行なう手法である。演奏表情生成規則とは、楽曲の旋律情報や楽曲の様々な特徴情報と演奏表情との関係を記述した規則である。ここでは、規則に基づく演奏表情生成手法に関する先行研究を概観する。

Sundberg らの研究 初期の演奏表情生成の代表的な研究として、Sundberg らによる分析による合成を用いた演奏表情の生成モデルの構築 [13, 15, 16] がある。この研究も規則に基づく手法を用いて演奏表情の生成を行なったものである。彼らの研究の本来の目的は、彼らが作成した演奏表情生成規則 [14, 17] の妥当性を評価することにある。彼らは音楽教育で一般的に知られている演奏ルールを分析して演奏表情生成規則を作成し、それを利用して演奏表情の自動生成を行なった。規則の組合せを色々と変更して演奏表情を生成し、その結果を個々の演奏表情生成規則の妥当性などの評価の材料として利用している。

Sundberg らの作成した演奏表情生成規則は主に以下の規則からなる。

- コード進行に関する規則

コードとは旋律の基調となる和音の組合せのことである。通常は一小節や半小節ごとに一つのコードを割り当てることができる。楽曲中のコードやその並び（コード進行）は楽曲の特徴を表す要素として非常に重要である。コード進行に関する規則は、コードの境界部分での演奏表情の変化に関する規則をまとめたものである。Sundberg らは、それぞれのコードの間の関係を表す数値として和音距離という概念を導入しており、この数値を利用して細かな演奏表情生成規則を作成している。

- フレーズの階層性に関する規則

フレーズとは、楽曲を構成している音楽的に意味のあるまとまりのことを指す。我々が音楽の一部を表す時に用いる主題や繰り返し、前奏、間奏などがフレーズである。一般に、楽曲は様々なフレーズを含んでいる。また、フレーズはより短い複数のサブフレーズを内包していることが多い。フレーズの階層性に関する規則は、曲全体をフレーズやさらに細かいサブフレーズに分割し、これらの境界部分で見られる音長などの変化をルール化したものである。一般に、フレーズの境界部分は他の部分に比べて演奏表情の変化が顕著になる傾向がある。例えば、大きなフレーズの間には短い休みをはさんだり、フレーズの終末部分にかけて徐々にテンポを遅くする、といったものが挙げられる。

- 調音距離に関する規則

調音距離に関する規則は、楽曲中の音高の推移と演奏表情との関係を記した規則である。楽曲の音高の推移も演奏表情を決定する上で重要な要素の一つである。例えば、音高と音の強さは密接な関係にあることが多い。Sundberg らは、音高の推移の特徴を表現するパラメータとして調音距離という概念を導入している。調音距離は音符同士の相対的な音高の変化や主音との差から導出する。

Sundberg らの演奏表情生成手法の手順を以下に簡単に示す。

1. 楽譜を計算機に入力
2. 音楽知識を入力
 - フレーズやサブフレーズの境界
 - コード
3. 調音距離を分析
4. 演奏表情生成規則を適用
5. 演奏表情を伴った演奏データを生成

入力として、演奏対象曲の楽譜情報や上記の演奏表情生成規則に用いられている音楽知識を付与する。フレーズやコードの情報をあらかじめ入力しているのは、これらは楽譜のみから判定するのが困難なためであると思われる。そして、対象曲のフレーズやコード、調音距離に対応する演奏表情生成規則を適用し、演奏表情の生成を行なっている。

彼らは、あらかじめ用意した演奏表情生成規則の中からいくつかの規則を選択した規則セットを作成し、その規則セットをシステムに付与している。この規則セットの組合せを様々に変更し、それぞれの規則セッ

トを用いて演奏表情の生成を行なっている。それぞれの生成された演奏を人間が評価し、良い演奏が得られた規則セットを分析した。その結果、良い演奏表情を生成する規則には以下に挙げる特徴が見られると主張している。

- 予期せぬイベントを強調する
- フレーズやサブフレーズなどの楽曲の構造を表現しようとする
- 強調したい部分を長く演奏する
- フレーズの最後でゆっくり演奏する

まとめ 本項では規則を利用した演奏表情生成手法の代表例として Sundberg らの研究を概観した。規則に基づく演奏表情生成手法はこの他にも、Bresin の研究 [8] や野池らの研究 [51]、平賀らの研究 [33, 37, 39] など数多く報告されている。これらの研究も、演奏表情生成規則を利用して演奏表情の生成を行なっている。演奏表情と関連付ける情報は楽曲の音高や音長など楽譜的な情報から認知的楽曲構造に至るまで様々であるが、基本的には、楽曲の旋律の特徴や楽曲構造と演奏表情とを関連付ける規則を人手で作成し、この規則を演奏対象曲に適用することで演奏表情の生成を行なうという点で共通している。これらの手法を用いるためには、演奏表情生成規則の他に、演奏曲の楽譜情報から、旋律の特徴情報や楽曲構造情報を取り出す知識が必要である。

2.2.2 演奏表情生成規則の学習

前項では、演奏表情生成規則を人手で作成し、これを利用して演奏表情を生成する手法について説明した。しかし、演奏表情の中には演奏者が無意識に行なっているものや、その根拠が判然としないものが多く、こういったものの規則を人手で作成するのは困難である。また、実際に演奏表情を生成するためには、テンポや音の強さをどう変化させるかだけでなく、どの程度変化させるかも決定しなければならないが、変化量のような定量的なパラメータの設定も、人手で行なうには適さない。

そこで、人手による規則では十分に記述できないこれらの演奏表情知識を取得するため、人工知能技術や学習アルゴリズムを用いて人間の演奏データから演奏表情生成規則を学習させる試みが行なわれている。

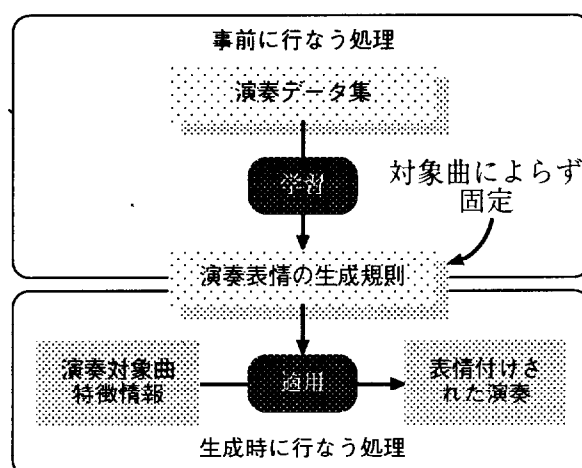


図 2.9: 演奏表情生成規則の学習を伴った演奏表情生成

Widmer の研究 その中でも代表的なものとして、Widmer による定性的認知モデルを利用した演奏表情の学習と生成がある。Widmer は説明に基づく学習 (Explanation-Based-Learning) を用いて演奏表情の生成規則を学習させ、さらにその規則を用いて演奏表情の生成を行なった [74, 75, 76]。彼は「演奏表情の役割は楽曲の構造を聴者に伝えることにある」という仮説をたてた。この仮説が正しいとすれば、楽曲構造に関する知識と聴者の演奏認知過程を組み合わせることで、演奏表情生成のモデルを構築できることになる。ここでの聴者の演奏認知過程とは、聴者の中にある演奏表情と楽曲構造とを関連付ける知識のことである。先にも述べたように、楽曲構造は既にある程度理論的に確立されているので、楽曲構造とその構造を適切に表現する演奏表情との関連づけの知識が得られれば、演奏表情生成が可能であると言えることを示している。

Widmer は、まず、演奏表情認知モデルの雛型を用意し、この雛型に楽譜情報と演奏表情の伴った演奏データを与えてその間の関係を学習させることで、人間の演奏表情認知モデルの構築を行なった [77, 78]。そして、この演奏表情認知モデルを逆に楽曲に適用することで、演奏表情の生成を行なった。

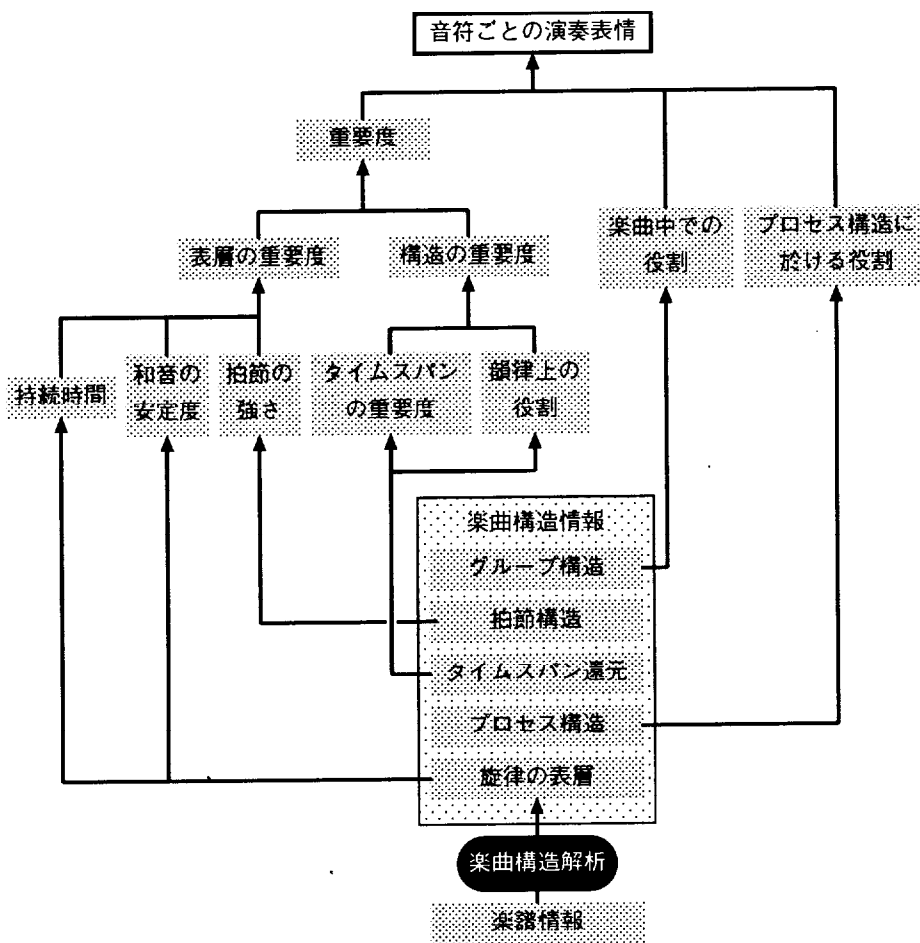


図 2.10: Widmer の構築した音楽認知モデル

Widmer の構築したモデルを 図 2.10 に示す。このモデルに導入されている音楽知識は、大きく分けると以下の二つである。

- 楽曲の構造に関する知識

これは、楽譜情報と楽曲構造との関連づけの知識である。この部分では、楽譜に書かれた旋律を分析し、その曲の構造情報を抽出する。楽曲構造情報やその抽出法は 2.1.4 項 で述べた楽曲構造の認知の理論に基づいている。このモデルでは以下の四つの情報を利用している。

- グループ構造
- 拍節構造
- タイムスパン還元
- プロセス構造

これらは音符の拍位置や旋律のグループ分け、楽曲中での音符の役割に関する知識である。この知識を楽曲に適用すると、フレーズの境界や、フレーズ内のそれぞれの音符の重要度など、旋律の楽曲構造の特徴属性が得られる。この知識は既知のものとして与えており、この知識に対しては学習は行わない。

- 楽曲の構造と演奏表情との関係付けの知識

楽曲構造と演奏表情との関連づけの知識は、クレッシェンドやデクレッシェンドなどの演奏表情の変化と楽曲構造との相関関係である。モデルの学習では、楽曲構造と演奏表情との関係に関する規則を学習する。

次に、この演奏表情認知モデルの学習の手順を以下に示す。

- 楽曲構造の分析

まず最初に、楽曲に楽曲構造に関する知識を適用し、楽曲の各部分での楽曲構造の特徴属性を抽出する。

- 演奏表情の特徴分析

演奏データの演奏表情を分析し、特定の演奏表情の変化が見られる箇所を抽出する。例えば、音の強さが徐々に強くなっている部分を見つけだし、その箇所をクレッシェンドが行なわれている場所として抽出する。

- 特徴箇所の楽曲構造の調査

演奏表情の変化が見られる箇所に共通してみられる楽曲構造の属性を調べる。また、同時に、注目している変化が見られない箇所についても共通する楽曲構造の属性を調べる。

- 傾向の分析

演奏表情の変化と関係する楽曲構造の属性を分析する。まず、演奏表情の特徴が見られる箇所に共通する楽曲構造の属性から、特徴が見られない部分にはない属性を選択する。そして、この属性を、注目した演奏表情の要因となる楽曲構造であるとする。例えば、クレッシェンドしている箇所の属性が、「フレーズの境界付近」と「重要度が高い」の 2 つだったとする。この中でクレッシェンドしていない場所では見られない属性がクレッシェンドと関係する属性である。もし、クレッシェンドしていない箇所でも「フレーズの境界付近」という属性が多く見られたとすると、クレッシェンドと関係する属性は「重要度が高い」の 1 つのみである。

この一連の手続きを様々な演奏表情の変化について行ない、その結果を演奏表情生成規則とする。

Widmer はこの手法を用いて演奏表情の学習と生成の実験を行なっている。まず、J. S. Bach のメヌエットから三曲を実験曲として選択し、合計 222 回分の演奏データを用意した。そして、それぞれの曲を前半

と後半に分割し、前半をテストセット、後半をトレーニングセットとした。トレーニングセットを利用して、テンポ変化と音の強さを学習させ、58 個の規則を学習した。学習した規則の内訳を表 2.2 に示す。この規則の中には他の研究で主張されている規則も多く含まれていた。

演奏表情	規則数
クレシェンド	14
デクレシェンド	15
リタルダンド	13
アツツェレランド	16

表 2.2: Widmer のシステムが学習した演奏表情生成規則の内訳

学習した規則集合を利用し、テストセットに対して演奏表情を生成させた。生成された演奏で聴取実験を行なったところ、よい評価が得られた。また、生成された演奏の中には、テストセットに対する人間の演奏と共通する特徴も見られた。規則集合から一部の規則を取り除いて演奏表情の生成を行なったところ、演奏表情が貧弱になった。Widmer は、これらの実験結果から、有効且つ必要十分な演奏表情生成規則が学習できた主張している。一方で、この規則集合をジャズの楽曲に適用したところ、あまりよい演奏表情が生成できなかった点を指摘している。

大阪大学の研究 大阪大学の青野らは重回帰分析を用いた演奏表情の学習と生成 [1, 2, 71] を提案している。また、この手法を利用して演奏表情の学習と生成を行なうシステム MIS を構築している。

重回帰分析とは、学習させたい情報群（目的変数）とそれに関連すると思われるパラメータ群（説明変数）との間に線形結合の関係が成り立つと仮定し、その間の関係を表す一次関数の結合係数を学習する手法である。これらの関係を式 (2.1)、式 (2.2) に示す。

$$\mathbf{y} \simeq \mathbf{X} \cdot \mathbf{b} + \boldsymbol{\varepsilon} \quad (2.1)$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \simeq \begin{pmatrix} x_{11} & x_{21} & \cdots & x_{m1} \\ x_{12} & x_{22} & \cdots & x_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{mn} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad (2.2)$$

$\mathbf{y} = (y_1 \cdots y_n)$ が目的変数、 $\mathbf{b} = (b_1 \cdots b_m)$ が説明変数である。この式の説明変数の線形結合 $\mathbf{X} \cdot \mathbf{b}$ と目的変数 \mathbf{y} との自乗誤差 $\sum_{i=1}^n \varepsilon_i^2$ を最小にする結合係数行列 \mathbf{X} の各項 x_{ij} を求めるアルゴリズムが重回帰分析である。重回帰分析はロバスト性が高く、学習結果の分析が容易であるといった利点がある。

MIS は楽曲の特徴情報を説明変数、演奏データから得られた演奏表情を目的変数として、その間の関係を表す結合係数を重回帰分析で求める。説明変数で利用している情報を以下に示す。

- 楽譜記号

MIS では楽譜上の演奏表情に関する記号を入力として与えている。楽譜記号は大別すると「局所アトリビュート」と「広域アトリビュート」の二種類からなる。局所アトリビュートとはアクセント記号やフォルテ、ピアノなど単一の音符に付帯する記号である。広域アトリビュートはスラーやクレッシェンドなど、連続する複数の音符に付帯する記号である。楽譜記号に関する説明変数は、いずれも記号がある場合は 1、記号がない場合は 0 を与える。

- 楽曲構造情報

MIS では Lerdaahl と Jackendoff が調性音楽の生成理論 [42] で提唱しているプロロンゲーション・リダクションに基づいて楽曲構造解析を行ない、その結果から得られる情報を入力として与えている。この理論では楽曲構造解析で旋律のグルーピングや、グループ内の最も重要な音符の決定を行ない、その結果を説明変数に使用している。なお、実験では楽曲構造分析は専門家が人手で行ない、その結果を入力として与えている。

- 状態の変化

楽譜記号の内、大局アトリビュートに関しては記号の有無以外に、記号の開始位置あるいは終了位置を表す説明変数を別途用意している。

- 説明変数の論理積

重回帰分析は説明変数と目的変数の関係を一次関数の係数として学習するため、非線形な現象を直接学習することはできない。この問題に対処するため、MIS では複数の説明変数の論理積も入力として与えている。入力に追加する論理積は、任意の説明変数の組合せにたいして学習と生成を試行し、生成された演奏表情と学習に用いた演奏表情の適合率の向上に貢献する組合せを順次選択する。

目的変数は、音の強さやテンポ値を直接利用するのではなく、隣接音符間でのこれらの数値の差分を用いている。実際の数値より、差分値の方が説明変数との間の関係が線形になりやすいと主張している。

青野らは、MIS を用いた演奏表情の学習と生成の実験を行なっている。Chopin 作曲の Walzer の提示部の前半の演奏データを用いて結合係数の学習を行ない、その結果を Walzer の提示部の後半や Mendelssohn の In Memoriam の前半に適用した。その結果生成された演奏表情は人間の演奏表情に近いものであり、演奏家からも「適切な演奏表情である」という評価が得られたとしている。また、結合係数を分析した結果、音楽学において言われている楽曲と演奏表情との関係を裏付ける規則が学習されていた。

関西大学の研究 関西大学の白川らはエキスパートシステムとニューラルネットワークを組み合わせ、音の強さに関する演奏表情生成規則を学習するシステム [53, 56] を構築している。また学習された演奏表情生成規則を利用し、演奏表情の生成を行なっている。

白川らの研究では音の強さの変化を規則で記述しやすいものとそうでないものにわけて考えている。演奏表情生成規則として記述しやすい演奏表情はエキスパートシステムに学習させ、微妙なパラメータの学習のように規則としての記述が難しいものはニューラルネットワークに学習させている。

エキスパートシステムで学習する情報は強弱記号に関する規則と、それ以外の規則の二種類である。強弱記号に関する規則は、楽譜上の強弱記号と音の強さとの関係に関する規則である。白川らの研究では、楽譜上に書かれた強弱記号を入力として与え、それに基づいて基本的な音の強さの変化を決定するため、強弱記号を特別に重要な情報として他の情報とは分けて扱っている。それ以外の規則は、強弱記号以外の楽譜情報と演奏表情との関係である。強弱記号が書かれていない部分での音の強さの変化や、強弱記号が同じ区間の中で音の強さの揺らぎを、楽譜情報と関連付けて記述したものがその他の規則である。ニューラルネットワークでは、エキスパートシステムでは学習できなかった音の強さの学習を行なっている。

白川らのシステムの演奏表情の学習手順を以下に示す (図 2.11)。

1. 強弱記号に関する学習

まず最初に、楽譜上の強弱記号が記されている部分の前後で、演奏データの音の強さがどのように変化しているかを分析している。そして、その変化量を強弱記号ごとに一次近似し、その結果を強弱記号に対する音の強さの変化量として学習している。学習は、フォルテやピアノなど相対的な音の強さを指定する記号と、クレシェンドなど音の強さの変化を表す記号に分類し、それぞれ別個に学習する。

2. その他の規則の学習

次に、強弱記号以外の楽譜情報に関する演奏表情生成規則の学習を行なう。ここでは以下の楽譜情報を利用する。

- 速度記号
- コード
- 繰り返し
- 拍子
- 曲中に於ける位置

まず、強弱記号に関する学習で得られた規則を利用して演奏表情を生成する。そして、生成された演奏表情と人間の演奏データとの差分をとる。この差分から隣接音符間の音の強さの変化を調べ、上記の楽譜情報との関連を学習させている。

3. ニューラルネットワークによる学習

ニューラルネットワークによる学習は一音ごとに行なう。入力層には、前後数個ずつの音符情報とエキスパートシステムで学習された規則に基づいて生成した演奏表情を与える。音符情報は音符ごとの、音長と音高の他に、楽譜に記載されている様々な記号も含まれている。ニューラルネットワークの教師信号には前後数個分の音符の人間による演奏の音の強さを与える。

実験はブルグミュラー作曲の「25 のやさしい練習曲」の演奏データから学習を行ない、その結果を利用して「進歩」の演奏表情を生成した。その結果、教師信号として利用した演奏者の演奏データに近い出力が得られたとしている。

白川らの手法は、ニューラルネットワークを導入することにより、規則による記述に向かない演奏表情生成規則も学習できるという利点がある。一方で、エキスパートシステムで学習する演奏表情が楽譜上の強弱記号に強く依存しているため、楽譜に強弱記号が記されていない場合や、強弱記号を重視しない演奏者の演奏データに対して学習を試みた場合に、演奏表情の学習や生成が適切に行なえるかどうか疑問である。

白川らは華やかな演奏と寂しそうな演奏のように、曲想が全く逆の演奏データを一緒に学習させると、平均的な演奏データの生成しかできなくなるという問題点を指摘している。その解決策として、曲想によって学習データを分類し、エキスパートシステムとニューラルネットワークの両方で、それぞれのグループごとに別の規則を学習するという手法の提案を行なっている。この手法によって寂しげな演奏の規則や華やかな演奏の規則が獲得できると主張している。しかし、この手法では生成される規則が別個のものとなるため、やや華やかな演奏のように、中間的な演奏表情の生成が行なえない可能性がある。

まとめ 本項では、学習に基づく演奏表情生成手法の中から、Widmer の演奏認知モデルを利用した手法、青野らの重回帰分析を利用した手法、白川らのエキスパートシステムとニューラルネットワークを組み合わせた手法について概観した。演奏表情生成を目的とした演奏表情生成規則の学習に関する研究としては、この他に遺伝的アルゴリズムを用いた手法 [80] などもある。また、演奏表情生成過程の解明を目的とした研究でも、演奏表情生成規則の自動獲得や、演奏表情生成過程のモデルの提案が行なわれている [9, 12, 24, 68, 69]。この他、規則に基づく手法でも、演奏表情生成規則の作成支援を目的とした演奏表情生成規則の分析や抽出の研究が行なわれている [52, 55]。

2.2.1 項と 2.2.2 項で述べた手法は、いずれも演奏表情に関する知識として、旋律の特徴や認知的楽曲構造と演奏表情とを関連付けた規則を利用しているという共通の特徴を持つ。つまり、演奏表情の知識を規則集合という形で抽象化して表現している。規則を用いた演奏表情生成は、演奏表情の生成の過程をコンパクトに表現できるという利点がある。一方で、生成規則に取り込むことができなかった知識が欠落して

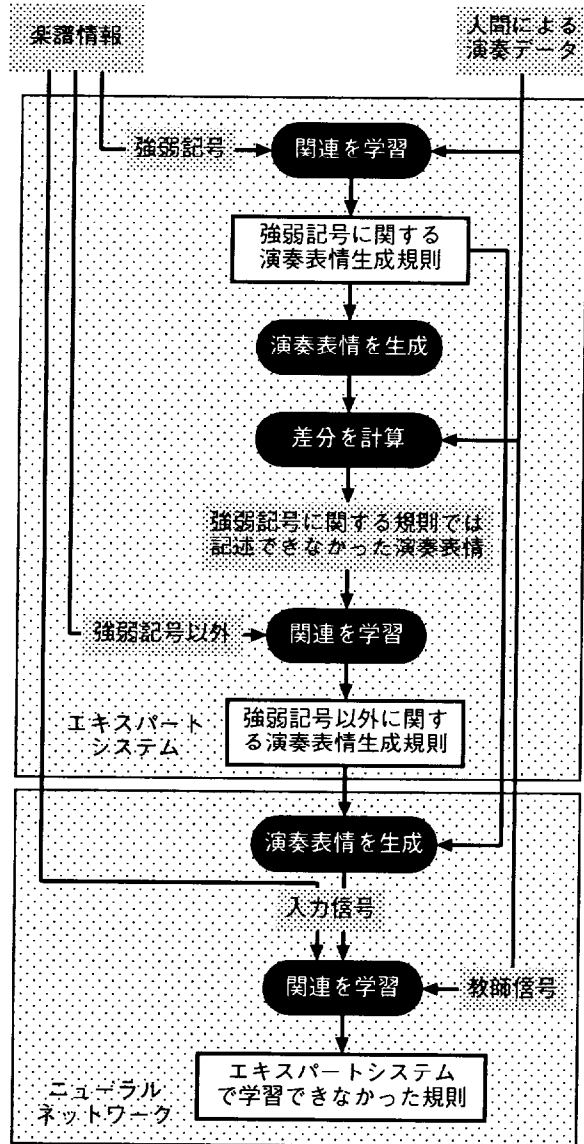


図 2.11: 白川らの手法での演奏表情生成規則の学習手順

しまうため、演奏表情を完全に再現することはできない。また、細かい部分まで学習しようとする、演奏表情生成規則や規則学習のためのモデルが複雑になってしまい、規則に基づく手法が本来持っている利点が損なわれてしまう。この欠点は、規則に基づく手法が本質的に抱えるものであり、規則に基づく手法を用いている限り避けて通ることができない問題である。

2.2.3 その他の手法

演奏表情生成規則に頼らない演奏表情の生成手法は多くはない。ここでは規則を利用しない手法として Arcos らのシステムを概説する。

Arcos らは事例に基づく推論 (Case-based Reasoning [40]) を応用した演奏表情生成手法を提案し、この手法を用いた演奏表情生成システム SaxEx [4] を実装している。事例に基づく推論とは、問題解決の手法の一種で、関連する問題の具体的な解決事例を応用し、新たな問題に対する解決を推論するという手法である。まず最初に、事例に基づく推論について簡単に説明する。

事例に基づく推論は以下に示す 4 つの処理からなる。これを 4R と呼ぶ。

- Retrieve

事例に基づく推論では過去に解いた問題の解決事例の集合（事例集）を問題の領域知識として利用する。Retrieve では、事例集の中から問題に類似した事例を検索する。まず最初に、与えられた問題を分析し、その問題の特徴を取得する。そして、事例集の中から、類似した特徴を持つ解決事例を検索する。さらに、検索された解決事例に対し、その有効性に基づいて順位づけを行なう。

- Reuse

Reuse では、Retrieve で得られた解決事例の分析を行なう。Retrieve で付けた順位の高い解決事例を選択し、その事例で、どうやって問題を解決したかを分析し、解決法を取得する。

- Revise

参考事例と与えられた問題が全く同じであるとは限らないので、参考事例の解決法を与えられた問題に直接適用できるとは限らない。Revise では、まず、Reuse で参考にした事例と、与えられた問題との相違点を分析する。そして、その相違に基づいて、事例から取得した解決法を与えられた問題に応用するためには、どのような修正や変更が必要かを検討する。これに基づいて解決法を修正し、与えられた問題の解決法を出力する。

- Retain

解決した問題は、新たな解決事例として事例集に追加する。この事例集への追加が Retain である。

事例に基づく推論は、規則に基づく手法のような領域知識の一般化を行わず、規則に基づく手法では学習データとして利用していた情報を全てそのままの形で利用する。従って、規則に基づく手法では学習による一般化で失われてしまう希少な事例も、事例に基づく手法では有効に活用することができる。また、学習のような前処理を殆んど行なわないので、新たな知識の追加にかかるコストが小さい。

Arcos らの手法では、過去に行なった演奏の記録を事例集とし、楽曲を問題として、与えられた問題に対する解決策（演奏表情）を生成する。この手法を実装した演奏表情生成システム SaxEx は、テナーサククスによるジャズバラードに演奏表情を付与するシステムである。その構成を図 2.12 に示す。

SaxEx は、入力として演奏表情が与えられていない演奏の音響データを入力として与えると、音響処理技術を用いてこのデータを加工し、演奏表情の与えられた音響データを出力する。前項以前で述べたシステムは、いずれも演奏のタイミングや音の強さを定量化した数値を入出力としており、演奏の音響データ

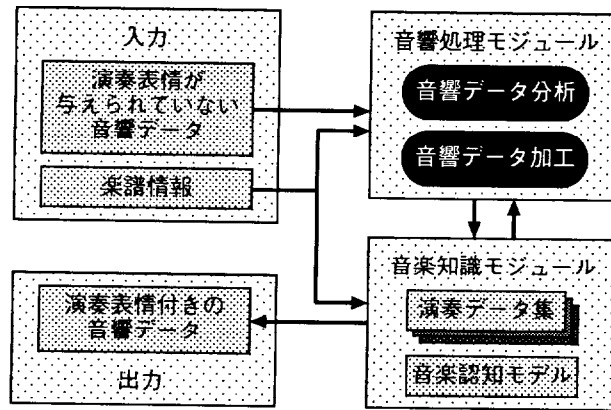


図 2.12: SaxEx の構成

を作成するためにはシンセサイザのような電子楽器を必要とするものであったが、SaxEX では演奏の音響データを直接出力することが可能である。これは、SaxEX が対象としているテナーサックスの楽器固有の奏法の中に、他の研究で対象としている鍵盤楽器に比べて数値での表現が難しいものも多く、また、演奏音の音響の複雑な変化を伴う表情付けが可能のためである。

SaxEx の演奏表情生成の手順を以下に説明する (図 2.13)。

1. 入力

演奏表情生成に利用する楽曲情報として、演奏曲の楽譜情報と演奏表情の与えられていない音響データを付与する。演奏曲の楽譜情報には、テナーサックスのパートのメロディーの情報の他に、コード進行など他のパートの情報も含まれているようである。

2. 分析

入力された音響データを信号処理し、各音の開始位置や終了位置を取得する。また、音楽認知理論に基づいて楽譜情報を解析し、楽曲構造情報を取り出す。音楽認知理論は、Lerdahl と Jackendoff の調性音楽の生成理論 [42] で提唱されている理論と Narmour の含意と実現のモデル [46] を利用している。この分析の結果に音高やコードなどの楽譜情報を合わせ、演奏曲の音符ごとに、音符の特徴を表す属性情報として記録している。SaxEx のデータ構造を図 2.14 に示す。analysis は楽曲構造分析から得られる情報である。prolong-structure と time-span-structure はそれぞれ調性音楽の生成理論の延長的還元とタイムスパン還元に対応する。IR-structure は Narmour の含意と実現のモデルである。score は楽譜情報である。melody はサックスパートの旋律の音符である。harmony はコードである。performance は演奏表情が与えられていない演奏データである。演奏データの音とサックスパートの楽譜の音符が一対一に対応している。

3. 検索

演奏曲の一音ごとに、演奏データ集の中から属性情報が類似している音符を検索する。そして属性情報の類似性から、検索された音符に順位付けする。

4. 応用

検索された音符の中から順位が一位の事例を取り出す。この事例を分析して演奏表情の変化量を抽出する。SaxEx では以下の 5 種類の演奏表情を取り扱う。

- 音の強さ
- テンポ変化

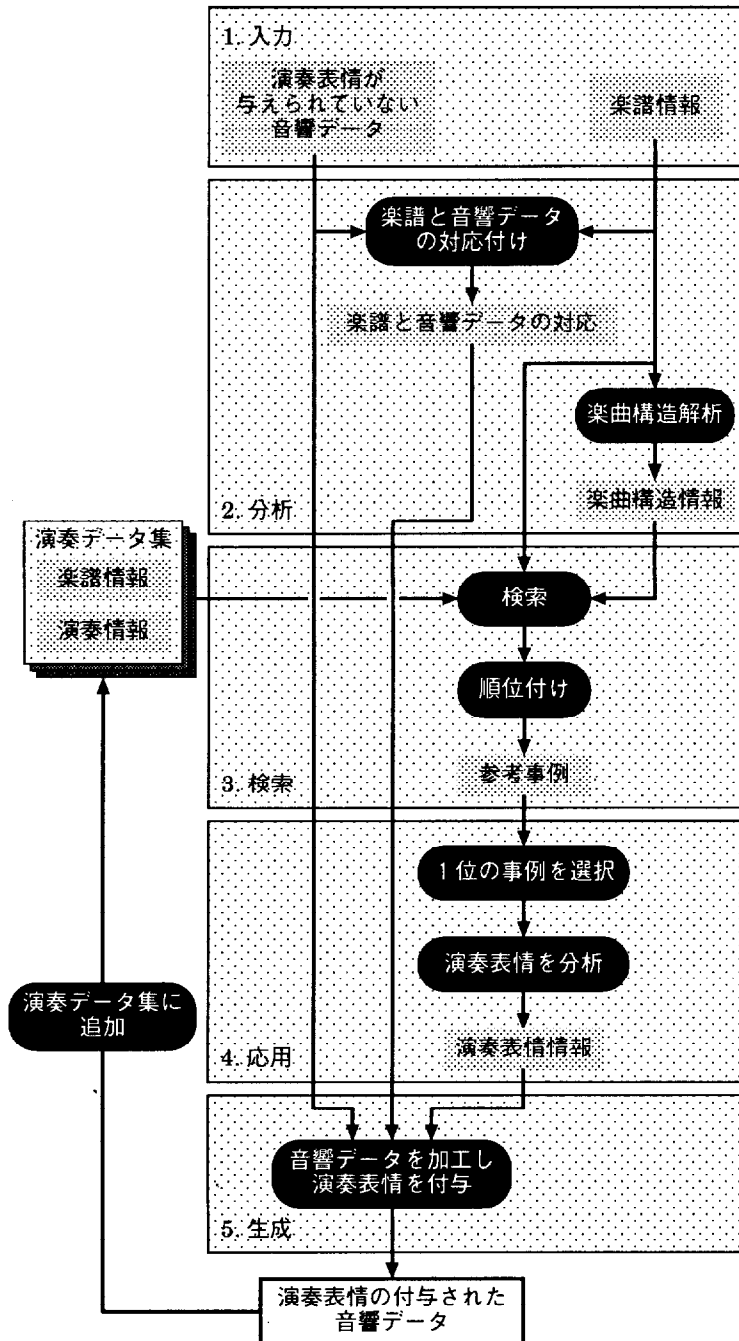


図 2.13: SaxEx の生成手順

- ビブラート
- 演奏長比
- アタック（サックス固有のもの）

いずれもその程度に応じて、数段階のラベルで表記する。そして、一位の事例の演奏表情を対象音の演奏表情とする。この手続きも、楽譜上の全ての音符に対して行なう。

なお、対象音の属性情報とここで取得した演奏表情との組合せを、新たな事例として事例集に追加する。

5. 生成

応用によって得られた演奏表情情報に基づいて、入力された音響データを加工し、演奏表情を付与した音響データを出力する。音響データの加工には独自の信号処理ツールを利用する。このツールは、テンポ変化やビブラート、音の強さなど上記の演奏表情を、演奏音の音高を変化させることなしに付与することが可能である。

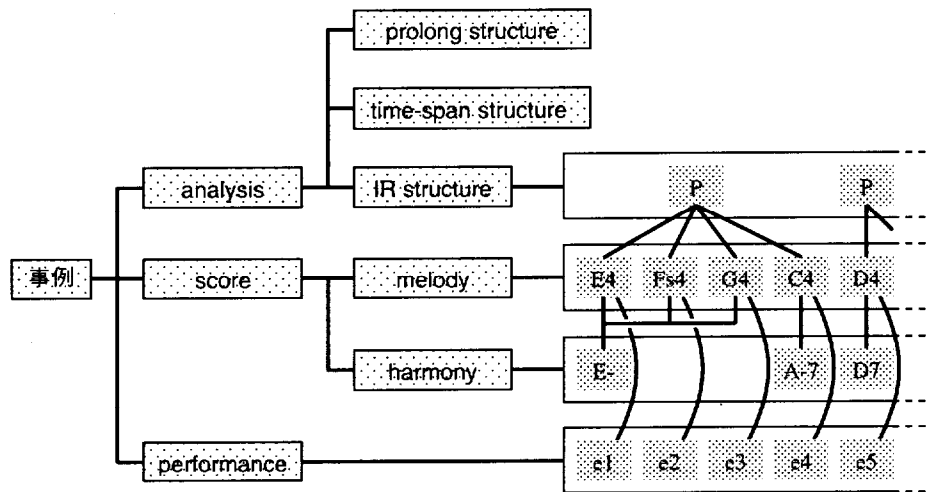


図 2.14: SaxEx のデータ構造

Arcos らは SaxEx を用いた演奏表情の生成実験を行なっている。実験には、ジャズバラード 4 曲を利用している。それぞれ、表情付けされた演奏と無表情な演奏の音響データを録音した。そのうち 1 曲から 20 音程度の短いフレーズを複数抜き出し、その音響データを事例集とした。まず、同じ楽曲の他の部分から抜き出した 20 音程度のフレーズを入力したところ、人間の演奏データとは多少の差異があるものの適切な演奏が得られた。また、3 曲から 50 音程度のフレーズを抜き出して事例集とし、残りの 1 曲から抜き出した 20 音程度のフレーズに対して演奏表情の生成を行なったところ、同様に適切な演奏表情が生成されたとしている。メロディーの推移のような大局的な演奏表情の変化も適切であったことから、音符の属性情報として導入している認知的知識が有効に機能していると主張している。

2.3 まとめ

本章では、演奏表情について解説し、演奏表情生成に関する先行研究を概観した。

演奏表情とは、音楽の演奏中に見られるテンポや音の強さの楽譜上からの芸術的逸脱のことである。演奏表情は音楽演奏を自然で味わい深いものに仕上げる上で必要不可欠である。テンポや音の強さの他にも

様々な要素があり、実際の演奏ではこれらの要素が複雑に絡みあって全体としての演奏表情を形成している。演奏表情の与え方自体は、これといって決まった型があるわけではなく、その与え方は演奏者に強く委ねられている。同じ楽曲の演奏でも、演奏者の趣向や意図、演奏時の雰囲気など楽曲外の要因（演奏状況）によって様々に変化する。

演奏表情をコンピュータに自動的に生成させる試みは数多く行なわれている。演奏表情はその生成に様々な音楽知識が必要となるため、コンピュータに自動的に生成させるのは容易ではない。演奏表情生成に関する研究は、大別すると以下の三種類に分類することができる。

- 規則に基づく手法
- 学習を用いた手法
- 演奏表情生成規則を利用しない手法

規則に基づく手法は、楽曲の特徴情報と演奏表情との間の関係を定義した規則を与え、それに基づいて演奏表情の生成を行なう手法である。この手法は、初期の演奏表情の自動生成で多く用いられていた。楽曲と演奏表情との関係は古くから演奏教育や演奏術などの分野で研究が行なわれていたため、基本的な規則であれば比較的容易に作成することが可能である。一方で、演奏表情の中には人手での分析し規則として記述することがが困難な要素もあるため、この方法では緻密な演奏表情の生成を行なうには限界がある。

学習に基づく手法は、楽曲と演奏表情との関係規則を人手で作成するのではなく、様々な認知モデルや人工知能理論を応用して、実際の演奏から学習する手法である。この手法は、規則に基づく手法が抱えていた規則作成の困難さを解消することが可能である。また、この手法を用いて学習された演奏表情生成規則を分析することにより、今まであまり知られていなかった人間の演奏時の認知行動を調査することも可能である。

演奏表情は、同じ楽曲に対して様々な演奏表情の与え方があることから、人間に近い演奏表情生成システムを構築するためには、演奏状況を考慮し単一の楽曲に多種多様な演奏表情の生成を行なうことが可能な手法が望まれる。しかし、演奏状況はパラメータが不特定で、なお且つ連続的に変化するため、演奏表情生成規則として記述したり学習させたりすることは容易ではない。従って、規則に基づく手法や学習を伴った手法では、多様な演奏表情の生成に関する研究はあまり行なわれていない。白川らのように、曲想ごとに別個の規則を学習させるという手法もあるが、あらかじめ規則の学習を行なった曲想の演奏表情しか生成できないし、中間的な演奏表情の生成も困難である。そもそも、これは演奏表情の種類の数だけ演奏表情生成規則の集合を用意するというものであり、多様な演奏表情が生成できる規則集合を作成するといった本質的な解決とはほど遠いものである。

演奏表情生成規則を利用しない手法を用いることで、これらの問題を解決できる可能性もあるが、これらの手法はまだ十分に研究されていない。

第3章 事例に基づく演奏表情生成

3.1 事例に基づく演奏表情生成

3.1.1 概念

まず最初に事例に基づく演奏表情生成手法 [58, 60] の概念について簡単に説明する。

演奏表情生成に関する先行研究の多くが演奏表情に関する知識を演奏表情生成規則という形で抽象化して保持していることは、2.2 節 で述べた通りである。規則に基づく手法は、学習を伴った演奏表情生成手法も含め、音楽演奏教育を受けた専門家の演奏表情生成過程モデルとみなすことができる。音楽演奏教育では、どういった場合にどのような演奏表情を与えるべきかについて様々なケーススタディが行なわれている。そして、演奏教育を受けた専門家はこのような教育を通して、楽曲の特徴や背景知識から最適な演奏表情がどのようなものであるかを導き出す自分なりの演奏表情決定規則を確立している。そして、専門家が演奏を行なう際には、演奏曲などの情報を自分の演奏表情決定規則にあてはめ、そこから導き出された演奏表情を対象曲に適用し、演奏を行なっている。

本研究では専門家モデルの対極として、音楽教育を受けていない素人の演奏表情生成過程モデルに注目する。演奏家の中には、専門的な音楽演奏教育を受けていないにも関わらず、非常に絶妙な演奏を行なう演奏家もいる。この中には、独学で演奏表情決定規則を確立した演奏家もいるが、演奏表情決定規則を持たず、専門家の演奏を模倣することによって表情豊かな演奏を行なう演奏家もいる。これらの演奏家は、演奏を行なう際に、演奏曲やそれに類似した楽曲に対する他の演奏者による演奏の記憶を思い起こしたり、あるいは新たにこれらの楽曲の演奏を聴取ことによって、他の演奏者が演奏曲にどのような演奏表情を与えているのかを調査する。そして、それを模倣して演奏を行なうことで、表情豊かな演奏を行なうのである。このアプローチを用いると、演奏表情生成規則を持っていなくても、演奏を記憶する能力と、演奏対象曲の演奏の聴取の経験さえあれば、少なくとも既知の楽曲に対して表情付けを行なうことが可能である。さらに、類似した楽曲の検索能力や応用力があれば、未知の曲に対しても表情付けを行なうことができる。

本研究で提案する事例に基づく演奏表情生成手法は、いわばこの素人の演奏家の模倣による演奏をモデル化したものである。事例に基づく演奏表情生成手法では、演奏表情に関する知識として、従来の演奏表情生成規則ではなく人間による演奏データ集を利用する。本手法では、前提として「旋律が類似していれば演奏表情が類似する」という仮定をする。これに基づいて、生成の過程では、入力として与えられた演奏曲に類似した楽曲の演奏データを演奏データ集の中から検索し、その中で演奏者が付与した演奏表情を分析する。そして、その演奏表情を模倣して演奏曲に適用することで演奏表情の生成を行なう。この流れを図 3.1 に示す。

規則に基づく手法では、演奏表情生成に必要な普遍的な知識として演奏表情生成規則を付与している。この知識は演奏表情に直接的に関わる知識であるため、規則に基づく手法では演奏表情に直結した知識が固定されていることになる。しかし、先に述べたように演奏表情の与え方は、演奏スタイルなどによって多様なバリエーションがあり演奏者ごとの個体差も大きい。従って演奏表情生成規則は演奏状況の変化に対する普遍性に乏しい。これが規則に基づく手法を用いて多様な演奏表情を生成することを困難にしている。

事例に基づく演奏表情生成手法では、普遍的な知識として旋律の検索や模倣の際に必要な旋律の類似性に関する知識を付与する。一方で、演奏表情に直接的に関わる知識である演奏データは、入力された対

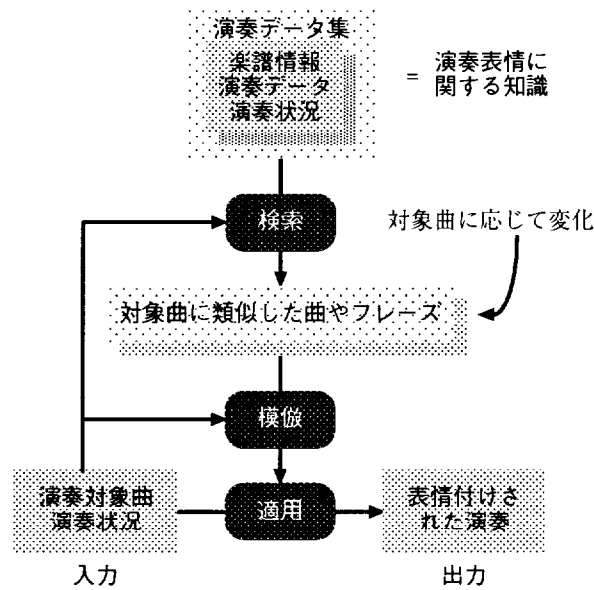


図 3.1: 事例に基づく演奏表情生成の概念

象曲や演奏状況を考慮して生成時に動的に取得するため可変である。このことから、規則に基づく手法に比べ柔軟性が高いといえる。また、普遍的な知識として固定されている旋律の類似性に関する知識も、演奏表情生成規則に比べて演奏スタイルや演奏者による差異が小さく、演奏状況の変化に対する汎用性が高い。したがって、本手法は規則に基づく手法に比べて異なる演奏状況に対応する能力に関して高いポテンシャルを持っている。

3.1.2 基本構成と手順

本項では、事例に基づく演奏表情生成手法の構成と機構を順に説明する。事例に基づく演奏表情の生成手法の基本構成を図 3.2 に示す。

事例に基づく演奏表情生成では演奏表情の知識として演奏データ集を利用する。この演奏データ集は人間が演奏したデータを集めたものである(図 3.3)。この中の個々の演奏データは以下の三種類の情報からなる。

- 演奏曲情報

演奏曲情報は主に演奏曲の楽譜情報で、音符情報、調性、拍子などから構成されている。音符情報は、楽曲中の各音符の音高、拍位置、音長で構成される。調性及び拍子は楽譜冒頭の調性記号と拍子記号によって示される情報である。

- 人間の演奏の記録

演奏の記録は演奏曲に対して人間が行なった演奏を記録したものである。これは、発音開始(打鍵)や発音停止(離鍵)などのイベントについて、時刻、音高、発音の強さなどの情報を時間順に列記したものである。

- 演奏時の演奏状況

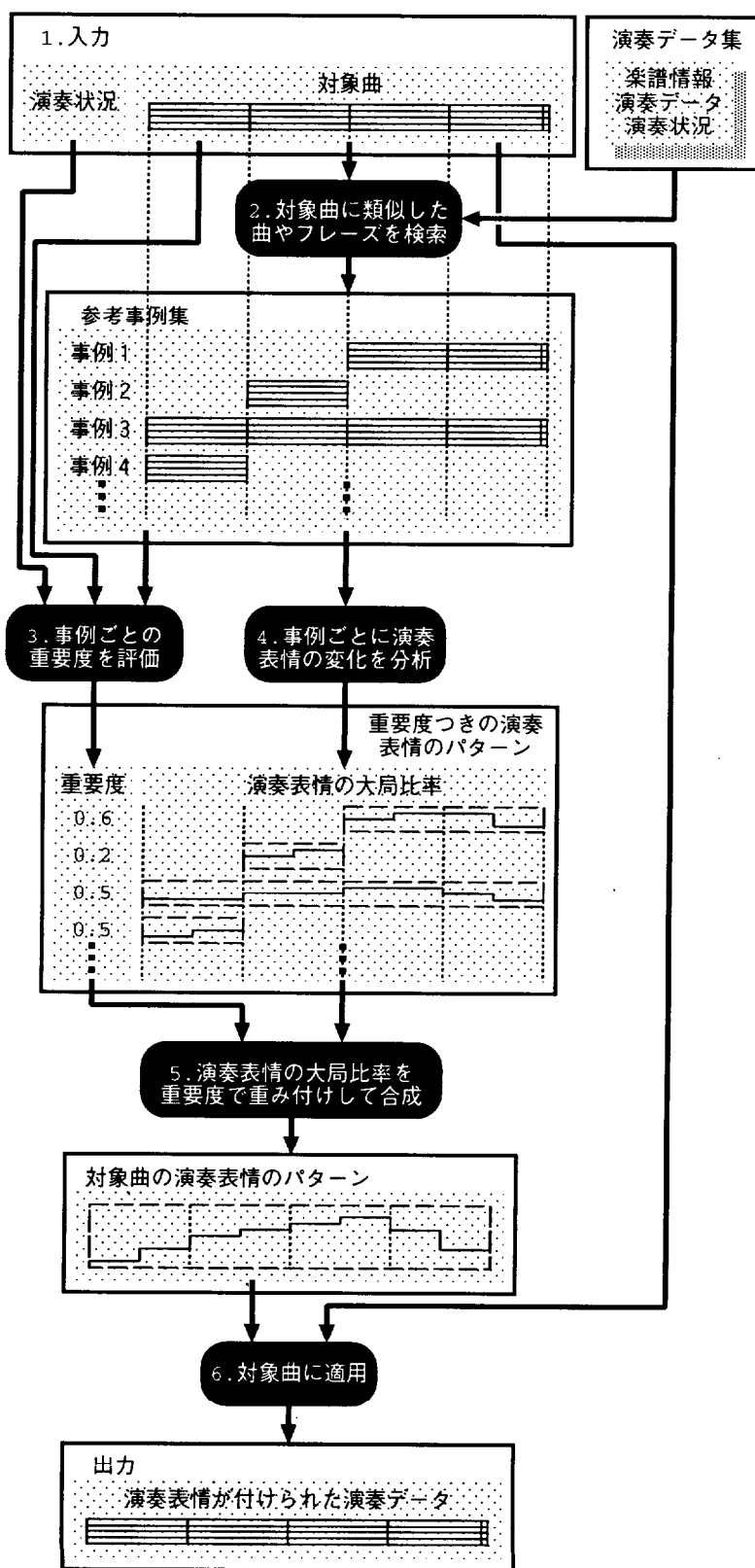


図 3.2: 事例に基づく演奏表情の生成手法の基本構成

演奏状況は、演奏データの演奏者や演奏データが収録された場所など、演奏に影響を与える演奏外の要因を表すパラメータである。

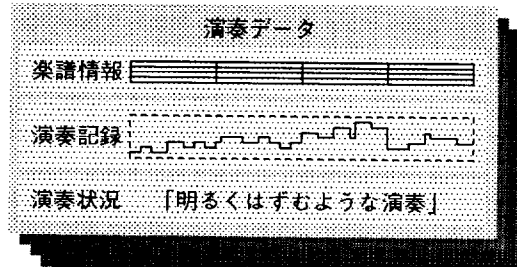


図 3.3: 演奏データ集の構成

演奏データ集には、楽曲全体の演奏データだけではなく、フレーズや小節など楽曲の一部分に対する演奏データも含まれている。

次に、事例に基づく演奏表情の生成手法の生成アルゴリズムを以下に示す。

1. 演奏対象曲の楽譜情報と演奏状況の設定を入力する
2. 演奏データ集から対象曲に類似した楽曲を検索し参考事例集を取得する
3. 以下の二つを並行して行なう
 - 個々の参考事例の重要度を評価する
 - 個々の参考事例の演奏表情を分析し演奏表情情報を取得する
4. 参考事例の演奏表情情報を合成し対象曲の演奏表情を作成する
5. 対象曲に演奏表情を適用し演奏データを出力する

入力 事例に基づく演奏表情生成手法では、入力として以下の二つの情報を必要とする（図 3.4）。

- 楽譜情報

楽曲情報は演奏対象曲に関する情報で、演奏データの楽曲情報と同様に楽譜情報などからなる。

- 演奏状況設定

演奏状況の設定は、生成しようとしている演奏データの演奏表情の特徴や雰囲気の規定する情報である。



図 3.4: 入力情報の構成

参考事例集の検索 システムは入力を受けると、まず最初に演奏データ集に対して演奏対象曲に類似した楽曲の検索を行なう(図 3.5)。この検索によって獲得される演奏対象曲に類似した楽曲の演奏データの集合を参考事例集と呼ぶ。参考事例の検索は演奏対象曲そのものだけではなく、演奏対象曲を構成するフレーズや小節などの部品についても行なう。このようにフレーズや小節など楽曲を構成する部分的な部品を旋律断片と呼ぶ。従って、参考事例集には楽曲だけではなく、フレーズや小節など様々な長さの演奏データが含まれる。これ以降の演奏表情生成の過程では、参考事例集に含まれる演奏データのみを利用する。なお、4.1 節に於いて、参考事例集の検索に関連した要素技術をとって事例の有効利用について議論する。

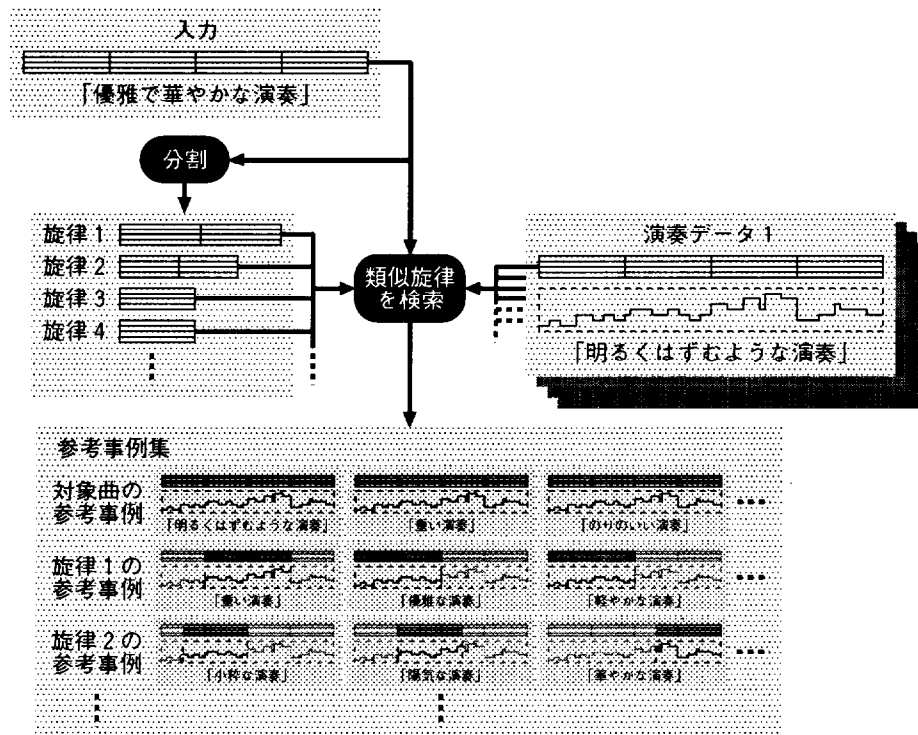


図 3.5: 類似旋律の検索

重要度の評価 演奏対象曲に対する参考事例集が得られたら、次の段階として、参考事例の重要度の評価と演奏表情の分析の二つの処理を並列して行なう(図 3.6)。

まず最初に、重要度の評価について概説する。この処理では、参考事例集の中の全ての演奏データについてその重要度の評価を行なう。重要度は対象曲の演奏表情を生成する際にその事例がどの程度参考になるかを表すスコアである。参考事例集に含まれる演奏データは、演奏対象曲にあまり似ていない楽曲のものから、非常によく似ている楽曲のものまで様々で、場合によっては演奏対象曲と全く同じ楽曲の演奏データが含まれている可能性もある。また、検索時には演奏データの演奏状況を考慮していないため、個々の演奏データの演奏状況はまちまちである。そこで、参考事例集に含まれる全ての事例について、演奏対象曲への演奏表情の生成を行なうに当たっての有用性を見定める必要がある。その指針となるのが事例の重要度である。

• 事例の重要度は以下の二つの情報から決定する。

- 演奏対象曲と事例の楽曲との類似性

- 入力された演奏状況設定と事例の演奏状況の類似性

いずれも入力と参考事例とが類似しているほど重要度は高くなる。小節やフレーズなどの旋律断片に対する事例の場合、事例の重要度は対象曲の旋律断片と事例の旋律断片の直接的な類似性だけでなく、それぞれの旋律断片の前後に連続する部分同士の類似性や周囲の旋律断片との関係も考慮して計算する。演奏状況についても、入力された演奏状況設定により近い演奏データに見られる演奏表情は、目標とする演奏表情に近い可能性が高いと考えられるので、演奏状況に近い事例にはより大きな重要度を与える。

なお、4.3 節で旋律の類似性の評価手法について、また、4.4 節で演奏状況の類似性の評価手法について、それぞれ提案を行なう。

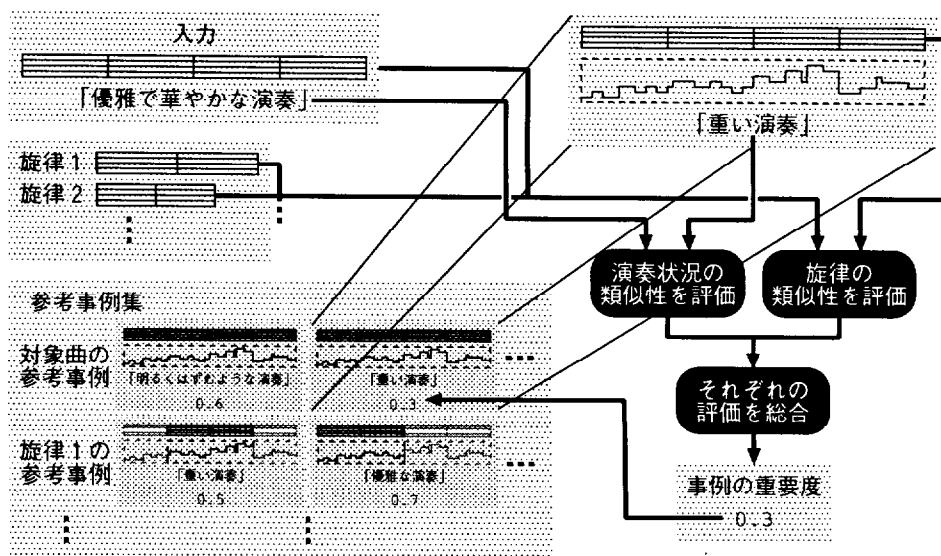


図 3.6: 重要度の評価

演奏表情の分析 重要度の評価と並行して、全ての参考事例に対して演奏表情の分析を行なう(図 3.7)。ここでは、参考事例に含まれる個々の演奏データで、人間がどのような演奏を行なったのかを取得するため、個々の演奏データから演奏表情情報を抽出する。事例に基づく演奏表情生成手法では、演奏表情情報として、テンポや音の強さの絶対的な数値(演奏表情値)ではなく、相対的な変化量の比率に変換したものを利用している。本手法で利用しているこの演奏表情の相対的な表記を演奏表情の大局比率と呼ぶ。本手法で演奏表情値ではなく演奏表情の大局比率を利用している主要な理由は、この次に行なう演奏表情の合成処理において、絶対的な数値より相対的な数値の方が扱いやすいことにある。したがって、演奏表情の分析処理では、個々の演奏データについて、まず最初に絶対的な演奏表情値の計算を行ない、そのあとでその演奏表情値を相対的な大局比率に変換する。なお、演奏表情の大局比率やその計算方法については4.2 節で詳しく述べる。

重要度の評価と演奏表情の分析の二つの処理が終了すると、重要度でスコア付けされた演奏表情の大局比率の集合が得られる。

演奏表情の合成 これまでの処理は、参考事例集の個々の事例を対象として、演奏表情の知識の抽出を目的とした処理である。これ以降では、これまでの処理で得られた演奏表情知識を対象曲に適用することを目的とした処理を行なう。

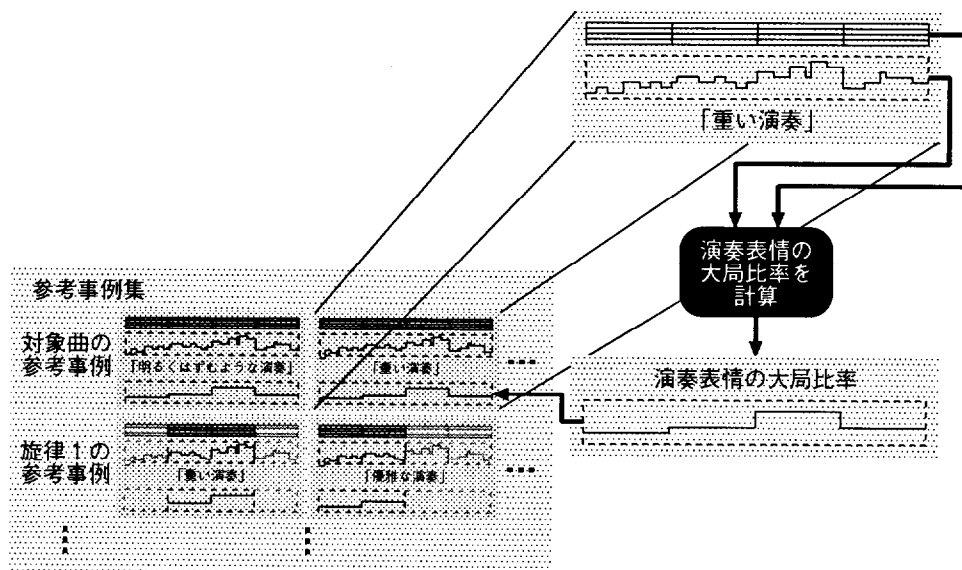


図 3.7: 演奏表情の分析

まず、前段階で取得された重要度でスコア付けされた演奏表情の大局比率を合成し対象曲の大局比率に変換する(図 3.8)。前段階で得られた重要度でスコア付けされた演奏表情の大局比率は、参考事例集に含まれる個々の演奏データに対するものである。参考事例集に含まれる事例は対象曲を構成する旋律に類似した旋律の演奏データであるが、その中には対象曲の同じ旋律に対する事例が複数含まれている可能性が高い。ここでは、その対象曲の同じ旋律に対する事例の演奏表情の大局比率をひとまとめにし、対象曲の旋律に対する演奏表情の大局比率を合成する。具体的には、同じ旋律に対する事例の大局比率を、その重要度に基づいて加重平均し、その結果を対象曲の旋律の大局比率とする。

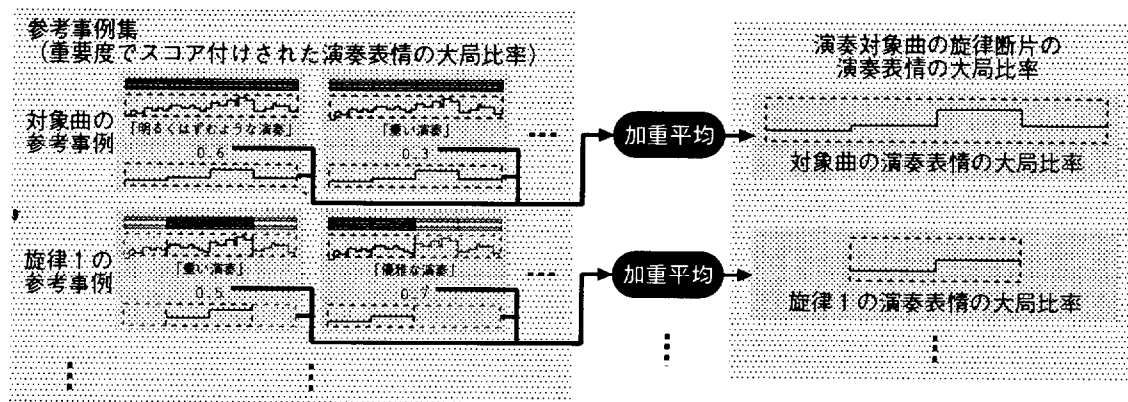


図 3.8: 演奏表情の合成

対象曲への適用 最後に、直前の処理で合成された演奏表情を対象曲に適用し、演奏データを出力する(図 3.9)。ここでは、大局比率で表記された演奏表情を絶対的な演奏表情値に変換する。演奏表情の大局比率から演奏表情値への逆変換については 4.2.2 項で詳説する。そして、その演奏表情値を演奏対象曲

の各音符に適用し、その結果を演奏データとして出力、もしくは演奏する。

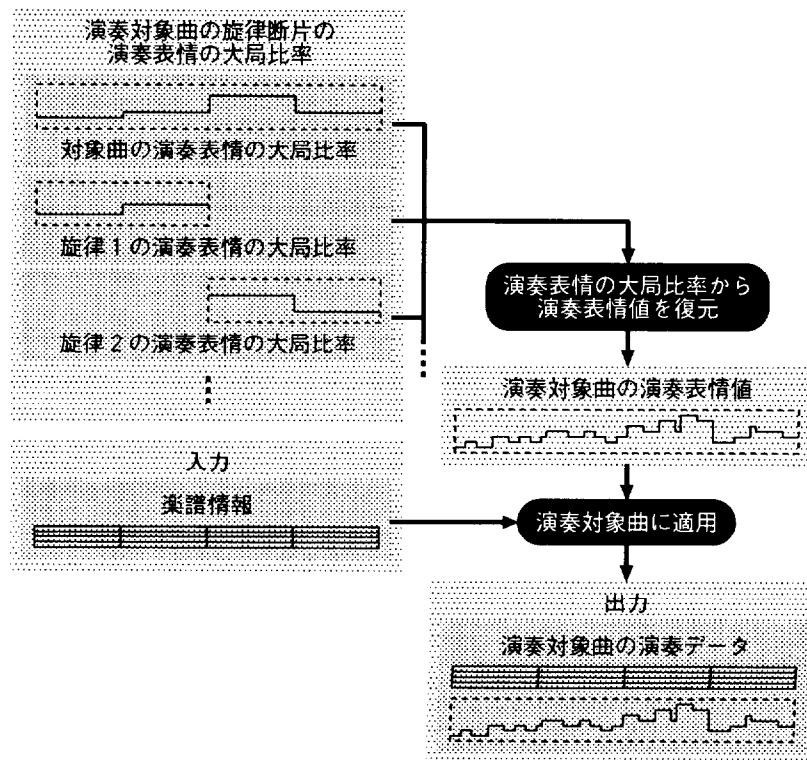


図 3.9: 対象曲への適用

3.2 特徴

3.2.1 利点

ここでは事例に基づく手法の利点について概説する。事例に基づく手法は概ね以下の3つの利点を持つ。

- 演奏状況を考慮した演奏表情の生成が容易

事例に基づく演奏表情の生成手法の最大の特徴は、演奏状況に応じた多様な演奏表情の生成が可能なことである。本手法では、重要度の評価の際に対象曲と参考事例との楽曲の類似性だけでなく、入力された演奏状況設定と参考事例の演奏状況の類似性も考慮している。従って同じ楽曲が入力された場合でも演奏状況の入力によって事例ごとの重要度の分布が変化する。演奏状況の近い事例により高い重要度が与えられることから、例えば、同じ楽曲に対する演奏データでもより入力された演奏状況設定に近い演奏状況を持つ演奏データの方が重要度が高くなる。演奏表情の生成では重要度が高い事例をより重視するため、最終的に生成される演奏表情は、演奏状況の近い事例に見られる演奏表情の特徴をより強く反映したものになる。これによって、入力された演奏状況の設定を反映して、様々な特徴を持った演奏表情の生成が可能となる。

従来の規則に基づく手法に演奏状況を導入するためには、演奏状況を演奏表情生成規則のパラメータに追加するか、演奏状況ごとに別々の演奏表情生成規則集合を用意するしかない。しかし、演奏状

況のパラメータは非常に多く、またそれぞれが連続的に変化するため、前者の手法では規則の作成や学習が困難であり現実的ではない。また、後者の方法は規則集合ごとに一つの演奏表情しか生成できないため、規則集合に与えられた演奏状況の中間に位置するような演奏表情の生成ができない。

- 認知的楽曲構造に関する知識が不要

第2章で述べたように、従来の演奏表情生成手法の中には「認知的楽曲構造が類似した旋律の演奏表情は類似する」という前提に基づくものが多い。この前提に基づいた手法の中には、認知的楽曲構造と演奏表情とを直接的に関連付けた規則を付与したり、あるいは、認知的楽曲構造を入力として与えるような演奏表情を出力するような関数のパラメータを学習させるなど、認知的楽曲構造に楽曲と演奏表情の仲介をする普遍的な知識として重要な役割を与えているものが少なくない。

しかし、認知的楽曲構造はその定義や解析方法が十分に明確であるとはいえず、また、楽曲構造の認知に個体差がある可能性も考えられることから、認知的楽曲構造に大きく依存することによって生じる弊害の危険性も否定出来ない。特に、これは演奏状況に応じた多種多様な演奏表情の生成において問題となる可能性が高い。演奏状況に応じた演奏表情生成では、様々な演奏家による様々な演奏表情が分析及び生成の対象となる可能性がある。しかし、全ての演奏家による演奏表情が、認知的楽曲構造との関係だけで説明できるとは断言できない。演奏家の中には、楽曲の認知過程があらかじめ用意した認知的楽曲構造とは異なる人がいる可能性もあるし、そもそも楽曲的認知構造への依存度が低い演奏表情を与える演奏家もいるかもしれない。従って、楽曲的認知構造に大きく依存した手法では、特定の種類の演奏状況下の演奏表情について適切な分析を行なうことができない可能性がある。

一方で、事例に基づく演奏表情生成手法は、生成の際に認知楽曲構造を利用する必要がない。本手法は、「類似した旋律の演奏表情は類似する」という前提に基づいている。そして、この前提に基づき、生成の過程では参考事例集（類似した旋律）から楽曲と演奏表情との直接的な関係を獲得しているため、生成過程に認知的楽曲構造は登場しない。勿論、参考事例集の検索や重要度の判定に要する知識に、認知的楽曲構造を導入することも可能であるが、それでも前述の従来手法のように認知的楽曲構造が無ければ演奏表情の生成が不可能であるということはない。従って、仮に認知的楽曲構造を導入したとしても、その抽出の確度による影響はそれほど大きくはない。

- 新たな知識の取得に要するコストが低い

本手法のもう一つの利点は、演奏表情に関する新たな知識の獲得が容易であるということである。学習を伴った規則に基づく手法では、新たな演奏データから取得した演奏表情を規則に反映させるためには、全ての演奏データを用いて規則の学習をもう一度やり直さなければならない。学習に利用する演奏データの数が膨大な場合や、新たな演奏データが頻繁に追加されるような状況では、学習に要する計算コストが大きな問題となる。一方で、事例に基づく演奏表情生成手法では、演奏データ集をそのまま演奏表情に関する知識として利用するため、新たな演奏データを取得した場合に、演奏データ集にそのデータを追加するだけで知識の獲得が終了する。従って、新しい知識（演奏データ）の追加の際の計算コストが少ない。

3.2.2 欠点

一方で本手法は以下の欠点を持つ。以降でそれぞれについて概説する。

- 生成時の計算コストが大きい

事例に基づく演奏表情生成手法は規則に基づく手法に比べて演奏表情生成時の計算コストが大きい。規則に基づく手法において演奏表情を生成する際に行なわれる処理は、大別すると演奏対象曲の分

析と演奏表情生成規則の適用の2つである。前者に要する時間は演奏対象曲の長さによって決まる。後者は対象曲の長さや演奏表情生成規則集合のサイズによって決まる。

一方、事例に基づく演奏表情生成手法は、参考事例の検索、参考事例の評価と分析、演奏対象曲への適用の3つである。それぞれの処理に要する時間は演奏対象曲の長さや演奏データ集のサイズ、演奏対象曲と参考事例集のサイズ、そして演奏対象曲のサイズに依存する。このうち、参考事例集のサイズは演奏データ集のサイズに比例するため、重要な要素は演奏データ集のサイズと対象曲の長さである。

演奏データ集のサイズと一般的な演奏表情生成規則集合のサイズとを比べると、演奏データ集の方が圧倒的に大きいので、一般的な規則に基づく手法に比べて生成時の計算コストが高い。

- 演奏表情の生成ができない場合がある

規則に基づく手法は、入力された楽曲に演奏表情生成規則を適用することによって演奏表情を適用することによって演奏表情の生成を行なう。汎用性が高い演奏表情規則を利用すれば、殆どどの楽曲に対して何らかの演奏表情を付与することが可能である。

一方で、本手法では演奏対象曲に類似した楽曲の演奏事例から演奏表情の付与例を取得し、それを演奏対象曲に応用することで演奏表情の生成を行なう。しかし、演奏対象曲に類似した楽曲が存在することが保証されていないため、十分な演奏事例が得られない楽曲に対して演奏表情の生成を行なうことができない可能性がある。

4.1節において、演奏事例が得られないという事態を減らすための要素技術を提案するが、これを用いても、演奏データ集と演奏対象曲の相性が悪い場合に演奏事例が得られない可能性があり、この欠点の解消には至らない。

- 生成過程が不透明

規則に基づく手法に比べると、本手法は演奏表情の生成の過程が不透明である。演奏表情の生成に関する研究の中には、演奏表情の生成ではなく、人間の演奏表情の生成過程の解明を目的とした研究もある。これは簡単に言えば上手な演奏家がどうやって優れた演奏を行っているのかを探求する研究である。これらは、人間に近い演奏表情を生成する演奏表情生成機構を探求することによって、人間が演奏表情を生成する際の知識行動のメカニズムの解明や、人間の音楽認知行動モデルの構築を行なうというものである。よって、これらの研究では、演奏表情の生成の過程で用いている中間的な情報や知識処理、さらにそれらの根拠が明確である必要がある。

一方で、事例に基づく演奏表情生成手法では、認知機構や知識行動のモデルを利用せず、ただ演奏データ中に見られる演奏表情を分析し対象曲に適用する。これは簡単に言えば上手な演奏家の演奏を模倣することで優れた演奏を行なうということである。このような戦略では、演奏表情生成の過程でどのような認知的な行動や知識が機能しているかがわからない。よって、事例に基づく演奏表情生成手法は人間の演奏表情の生成の分析などを目的とした分析的研究には適さない。

3.2.3 関連研究との相違

本手法は事例に基づく手法であるという点で Arcos らによる事例に基づく推論を用いた演奏表情生成 [4] に比較的近い。本項では主に Arcos らの事例に基づく推論を利用した演奏表情生成と事例に基づく演奏表情生成との相違点について説明する。

主要な相違点を以下に示す。

- 事例検索の単位

Arcos らの手法では、類似事例の検索や事例の分析、演奏表情の付与など、一連の処理を全て一音単位で行なっている。従って、Arcos らの手法では、音符ごとの事例から音符間の演奏表情の繋がりや全体的なバランスなどに関する知識を利用していない。勿論、事例の中の演奏表情は全体的なバランスなどを考慮したものであるため、事例から取得される一音ごとの演奏表情値もこれらの要素を含んでいるが、その値が的確であるかどうかには疑問がある。一方で、本手法ではこれらの処理は全て楽曲全体や楽曲を構成する旋律を単位として行なっているため、音符ごとの繋がりや全体的なバランスも考慮した知識の取得が容易である。なお、この問題については 4.1.1 項で詳説する。

- 事例の適用方法

Arcos らの手法では、演奏対象曲に含まれる音符ごとに、類似音符の検索の結果得られた参考事例の中から最も類似している事例一つを選択し、そこで行なわれている演奏表情をそのまま適用している。従って、一位になった事例が演奏ミスなどによって正しく演奏されていなかった場合、それがそのまま対象曲に適用されてしまう。事例に基づく演奏表情生成手法では、検索の結果得られた参考事例集の個々の事例に類似性を考慮して重み付けを行ない、そこで行なわれている演奏表情をこの重みで加重平均したものを対象曲に適用している。従って、最も類似している事例に演奏ミスが含まれていたとしても、他に演奏ミスがない参考事例がいくつかあれば、演奏ミスの影響が軽減された演奏表情の生成が可能である。

- 演奏表情の多様性

Arcos らの手法は現状では演奏状況に類する情報については考慮していない。従って、Arcos らのシステムは現状では同じ楽曲に対して多種多様な演奏表情の生成を行なうことはできない。

第4章 要素技術

本章では前章で述べた事例に基づく演奏表情生成手法を実現するにあたって必要となる要素技術について述べる。本章では主に以下の4点について述べる。

- 演奏データの有効利用に関する技術
- 演奏表情の記述，計算法
- 旋律の類似性の評価手法
- 演奏状況の記述と評価

演奏データの有効利用に関する技術は，参考事例の不足の解消を目的としたものである。事例に基づく手法では有効な参考事例が検索できないと問題を解決できない。この点については前章で紹介した事例に基づく演奏表情生成の場合も同様で，参考となる演奏データが得られないと適切な演奏表情が生成できない。そこで，限られた演奏データ集を有効利用し，可能な限り参考事例の不足が発生しないような方策を考える必要がある。4.1節では，参考事例の不足の問題を軽減するための技術として，対象曲を分割し参考事例を検索する手法について述べる。

演奏表情生成に限らず，計算機上で演奏表情を取り扱うためには演奏表情を数値として表現する必要がある。演奏表情の表現方法としては音符ごとにテンポや音の強さの数値を記録するのが標準的であるが，4.1節で紹介する参考事例の有効利用技術を導入すると，音符単位の記述では不都合が生じる。そこで，事例に基づく演奏表情生成手法に適した演奏表情の表現方法について4.2節で提案する。またこの演奏表情の表現方法を用いた演奏表情の分析や合成などの計算についても述べる。

事例に基づく演奏表情の生成手法では，参考事例の検索や事例ごとの重要度の計算の際に，旋律同士の類似性を評価する手法が必要となる。とりわけ，旋律の長さによらず類似性が評価できることや，旋律の全体的な特徴の類似性が評価できることが必要である。4.3節では，これらの要件を満たすような旋律の類似性の評価手法を提案する。

事例に基づく演奏表情生成の大きな特徴の一つとして，演奏状況に応じた演奏表情の生成が可能であるという点がある。演奏状況に応じた演奏表情の生成を行なうためには，演奏状況そのものを数値的に表現する必要がある。また，事例の重要度の評価の際には事例の演奏状況同士の近さを評価する必要がある。4.4節では，演奏状況の計算機上での表現方法と，この表現方法を利用した演奏状況の類似性の評価手法について提案する。

4.1 事例の有効利用

一般に事例に基づく手法が潜在的に抱える問題として参考事例の不足がある。事例に基づく手法では，データ集の中から対象に適用可能な事例を検索し，それを分析することによって問題の解決に必要な情報を獲得する。従って検索で得られる参考事例が少ないと必要な情報が得られないため，最終的な結果も有効なものを得られない。この問題は第3章で提案した事例に基づく演奏表情手法でも起こり得る問題である。本節では，事例に基づく演奏表情手法の参考事例の不足の問題を軽減する技術 [63] について述べる。

4.1.1 事例の不足とその影響

本手法において最も理想的な参考事例は対象曲そのものの演奏データである。対象曲と同じ楽曲の演奏データで演奏状況も入力と同じものがあれば、そのデータをそのまま出力データとすることができる。しかしながら、演奏データ集の中にこのようなデータが含まれることは期待できない。また、完全に同じ楽曲ではないにしても曲全体が似ているような楽曲があればかなり有用な事例となり得るが、いずれにせよ楽曲の長さや構成が殆んど一致している必要があるため、こういった事例も見つかる可能性が低い。となると、曲全体を検索単位として類似している曲を探すだけでは、十分な量の参考事例が得られる可能性が低く現実的ではない。

この問題を解決するには手法は検索の単位を小さくするのが有効である。この手法は Arcos らの事例に基づく推論による演奏表情生成 [4] でも利用されている。Arcos らは事例の検索を音符単位で行なうことで事例不足の問題を回避している。これは、対象曲を構成する音符ひとつひとつについて、事例集の中から音符長や音高、認知構造などの特徴が類似している音符を検索し、その1音に与えられている演奏表情を利用するというものである (2.2.3 項参照)。検索の単位が1音であるため、演奏データの楽曲数がそれほど多くなくても十分な量の事例が得られるし、また、楽曲そのもの同士を比較するのに比べて考慮すべき特徴情報が少なく済むため、事例不足の問題の軽減には有効である。

しかしながら、音符単位の検索では楽曲の流れを適切に考慮した演奏表情ができない可能性がある。演奏表情はその音の前後に連続する音符やフレーズの流れによって影響を受ける。従って、音符一個単位で決定できるものではなく、前後につながる音符や旋律を考慮する必要がある。Arcos らの手法では音符の特徴を分析する音楽知識モデルに、Narmour の含意と実現のモデル [46] や Lehdahl と Jackendoff による調性音楽の生成理論 [42] などの音楽認知に関する知識を導入している。これらの音楽認知に関する知識は音符単独ではなく、前後に接続する音符との関係から得られる特徴情報であるため、これらを導入することにより音符の前後の旋律の特徴も考慮することが可能である。しかしながら、これらの認知的音楽知識は、数小節程度の比較的短い範囲の旋律情報から得られる知識であるため、その範囲を越えるような長さの範囲の流れによって生じる演奏表情 (例えば、曲全体に対するフレーズごとの音の強さのバランスなど) に対しては有効に機能しない。

4.1.2 旋律断片を単位とする検索

本手法でも、事例不足の問題を解決するために、対象曲を分割し検索単位を小さくする手法をとる。しかし、Arcos らの手法のような音符単位の検索は前項で述べたような問題があるため、本手法では検索単位を音符ではなく旋律とする。検索時に、対象曲や演奏データ集のデータを様々な長さの旋律に分割し、それぞれの旋律について参考事例の検索を行なう。この分割した旋律を旋律断片と呼ぶ (図 4.1)。

ここでは、楽曲を旋律断片に分割した際の事例不足の問題に対する効用について概説する。まず、対象曲もしくは対象曲の旋律断片 s に対して、演奏データ集の中から参考事例が発見できる可能性 p ($0 \leq p \leq 1$) について考える。参考事例の発見確率 p は、事例の総数 n と平均的な旋律類似確率 q ($0 \leq q \leq 1$) を用いると式 (4.1) で表される。

$$\begin{aligned} p &= 1 - (1 - q)^n \\ &= 1 - \sum_{i=0}^{n-1} (-1)^i \cdot {}_n C_i \cdot q^i \\ &= - \sum_{i=1}^n (-1)^i \cdot {}_n C_i \cdot q^i \end{aligned} \quad (4.1)$$

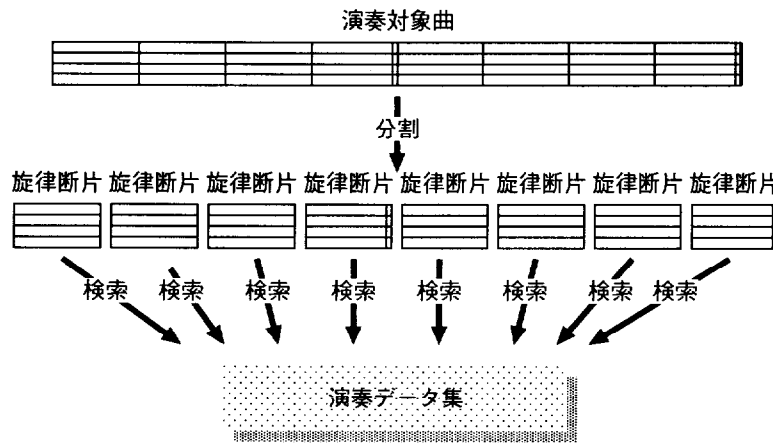


図 4.1: 演奏対象曲を旋律断片に分割してから検索

旋律類似確率 q は旋律 s と演奏データ集の中の旋律とを比較した際に、その旋律が参考事例として有効であると判断される確率である。この確率は旋律検索の際に使用する類似性の判定手法によって決定する。検索単位を曲全体とした場合 n は参考データ集の中の演奏データの総数となる。

検索単位を短くすることによる事例不足の問題に対する効用は主に以下の 2 点である。

- 事例の総数の増化
- 旋律の類似確率の向上

まず、一点目の事例の総数について考える。対象曲を仮に m 個の断片に分割した場合、検索単位の平均長は曲全体を検索単位とした場合に比べて平均で $\frac{1}{m}$ に減少する。対象曲のみを分割したのでは、演奏データ集の事例との長さが異なってしまい、事例として適切ではなくなるので、演奏データ集の個々の演奏データも平均の長さが対象曲の検索単位に近くなるように分割する必要がある。よって、分割によって対象曲の事例の総数が増える (図 4.2)。対象曲を m 分割した時の事例の総数 n_m と曲全体を検索単位とした場合の総数 n_1 は式 (4.2) 式に示す関係にある。

$$n_m = \frac{l_a \cdot m}{l_s} \cdot n_1 \quad (4.2)$$

l_a は演奏データ集の中の演奏データの平均長、 l_s は対象曲の長さをあらわす。演奏データの平均長と対象曲の長さが同じであると仮定すれば、事例の総数は単純に m 倍になると考えて良い。

次に、旋律の類似確率について考える。旋律の類似確率は、旋律の類似性の判断手法によって変化するため、一般化するのは容易ではない。ここでは、比較する旋律同士を全体にわたって一定の基準に基づいて比較したと仮定する。旋律が参考事例として有効であるとするためには、旋律が局所的に類似しているだけでは不十分で、旋律全体に渡って一様に対象と類似していることが重要である。

ここで図 4.3 に示すような場合について考える。この例に示す対象曲は前半と後半の二つの部分に分割できるものとする。さて、演奏データ集の中に次のような演奏データがあったとする。

- 対象曲と同様に二つに分割することができる
- 前半部分は全体的に類似性が高い
- 後半部分は類似性が低い

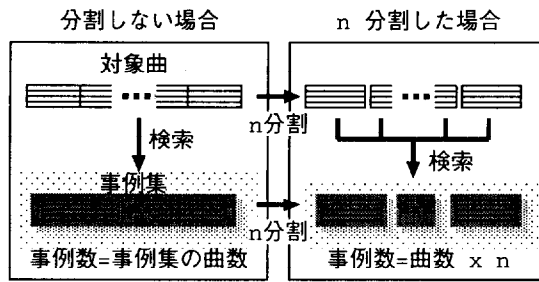


図 4.2: 旋律の分割による事例の総数の増化

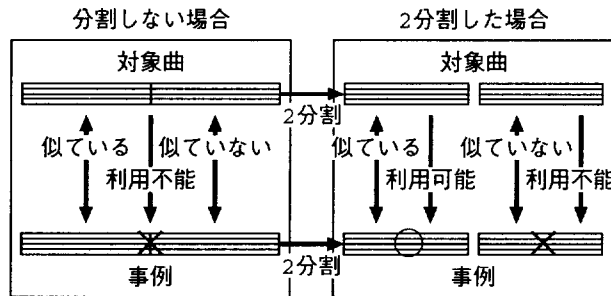


図 4.3: 旋律の分割による旋律の類似確率の向上

検索単位を対象曲全体とした場合、この事例は後半部分同士が似ていないということから、旋律全体に渡って一様に類似しているという先述の要件を満たしていないため、図 4.3 の左側に示すように参考事例としては不適切である。次に、この対象曲を前半部分と後半部分に二分分割して検索したとする。この場合、対象曲との検索単位を揃えるために、先の演奏データも前半と後半に二分分割することになる。すると、後半部分どうしはあまり類似していないので相変わらず参考事例としての利用はできないが、前半部分同士は全体的に類似性が高いため、対象曲の参考事例として利用することが可能である。したがって、この例では曲全体を検索単位とした場合には全く利用できなかった演奏データが、旋律を分割することによって利用可能になる。

事例の総数の増化の場合と同様に、楽曲を m 個の旋律断片に分割した場合について考える。検索単位を楽曲全体とした場合、ある事例が対象曲の参考事例として利用可能であるためには、その旋律を構成する m 個の旋律断片全てが互いに類似している必要がある。したがって、検索単位が曲全体のときの類似確率 q_1 と m 分割した場合の個々の断片についての類似確率 q_m の間には式 (4.3) に示す関係がある。

$$q_1 = q_m^m \tag{4.3}$$

q_m は 1 以下の数であるため、 q_m^m は q_m よりはるかに小さい値となる。

事例の総数の増化式 (4.2) と旋律の類似確率の向上式 (4.3) を参考事例の発見確率に適用すると、曲全体で検索した場合の発見確率 p_1 および、 m 分割した場合の発見確率 p_m はそれぞれ式 (4.4) と式 (4.5) になる。

$$p_1 = - \sum_{i=1}^n (-1)^i \cdot {}_n C_i \cdot q_1^i \tag{4.4}$$

$$\begin{aligned}
p_m &= -\sum_{i=1}^{m \cdot n} (-1)^i \cdot {}_n C_i \cdot q_m^i \\
&= -\sum_{i=1}^{m \cdot n} (-1)^i \cdot {}_n C_i \cdot q_1^{\frac{i}{m}}
\end{aligned} \tag{4.5}$$

q が十分に小さい数であるとする、 $\sum_{i=1}^n (-1)^i \cdot {}_n C_i \cdot q^i \simeq -n \cdot q$ と近似することができるので、これにより q_1 と q_m の間には概ね式 (4.6) に示す関係がなりたつ。

$$\begin{aligned}
p_1 &\simeq n \cdot q_1 \\
p_m &\simeq m \cdot n \cdot q_1^{\frac{1}{m}} \\
&\simeq m \cdot p_1^{\frac{1}{m}}
\end{aligned} \tag{4.6}$$

p_1 は 1 より小さい値であることから、 p_m は p_1 に対して指数オーダの m 倍という十分に大きな値になる。

4.1.3 検索単位が大きい事例の取り扱い

本論文で提案する手法は、認知的楽曲構造を導入しなくても前後に連なる旋律を考慮した演奏表情が取得できる。また、Arcos らの手法では弱かった認知的楽曲構造では考慮していない長い範囲の旋律の関係を考慮することも可能である。しかしながら、前項で述べた手法は曲全体など大きな旋律断片に対しては有効に機能しない。勿論、小さい旋律断片の参考事例を組み合わせることによって、演奏表情の生成を行なうことも可能ではあるが、全体的なバランスやより大きい範囲での演奏表情のつながりのような長い範囲に影響を与える演奏表情を考えると、大きい旋律断片に対する事例も利用すべきである。本項では、このような大きい旋律断片の参考事例に対する事例不足の問題を軽減する手法について述べる。

式 (4.6) に示す通り、参考事例となる旋律が見つかる確率は、旋律断片の分割数、もしくは検索対象となる旋律断片の平均長に大きく影響する。従って、旋律断片の長さが長くなるにしたがって楽曲の分割数 m は 1 に近くなるため、分割による効用が小さくなる。

楽曲の分割数 m が 1 となる曲全体やそれに近い長さの旋律断片に対しても、参考事例の発見確率 p を向上させるためには、旋律の類似確率 q を高くするしかない。 q は旋律の類似性の評価方法に依存することから、 q を向上させるためには旋律の類似性の判定条件について検討する必要がある。

前項では、旋律の類似性を、旋律断片の長さによらず、比較する旋律同士を全体にわたって一定の基準に基づいて比較することを前提としていた。従って、旋律断片が長くなれば類似性のスコアはより短い旋律断片の類似性の積になるため、式 (4.3) に示すように旋律断片の長さが長くなるに従って、指数的に旋律断片の類似確率 q が低下する。これを回避するためには、旋律の長さによって類似性の評価基準が変化するような手法を導入する必要がある。

事例に基づく演奏表情生成手法では、参考事例を検索する際に旋律断片の大小に関わらず旋律断片が持つ全体的な特徴同士のみを比較し、その結果から旋律断片の類似性を評価する。この手法を用いると、検索される旋律断片の子細な部分に関する類似性が保証されなくなるが、その一方で、旋律断片の分割数 m に関わらず旋律同士の類似確立 q をほぼ一定の高い値にすることが可能である。なお、この手法の詳細は 4.3 節で述べる。

旋律の全体的な特徴の比較を用いて収集した参考事例は、先に述べたように旋律の類似確率の向上の代償として、子細な部分の類似性が保証されなくなるという問題がある。従って、このような参考事例を全面的に参考にして演奏表情の生成を行なうのは適切ではない。全体的な演奏表情の傾向はともかく、子細な部分に与えられた演奏表情は演奏対象曲にそのまま適用するには適さない可能性が高い。

ここで、今一度、旋律の全体的な特徴の比較を導入した経緯を整理すると、以下のようになる。

- 旋律の分割による効用は長さの長い旋律断片に対しては小さい
- 演奏表情の全体的なバランスなどのことを考えると、長さの長い旋律断片の参考事例も必要である

つまり、長い旋律断片の子細な部分で生じている演奏表情については、無理に長い旋律断片の参考事例から抽出しなくても、より短い旋律断片の参考事例から取得することが出来る。そして、長さが短い旋律断片は、前項で述べた分割による効用が大きいため、演奏表情の取得に関してはさほど心配ない。このことから、より長い参考事例からは、短い断片同士の間全体的な演奏表情のバランスや繋がりなど、旋律断片全体に渡って生じている演奏表情を抽出するだけで十分である（図 4.4）。

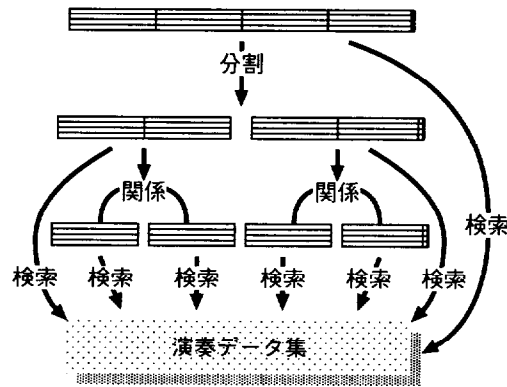


図 4.4: 参考事例から大局的な演奏表情のみを取得

そこで、事例に基づく演奏表情生成では、個々の参考事例からは全体的な演奏表情のバランスのみを抽出し、より短い旋律断片から抽出された子細な部分の演奏表情と組み合わせることで最終的な演奏表情の生成を行なう。そして、全体的な演奏表情のみの獲得をするために、演奏表情の大局比率という概念を導入している。これについては 4.2 節 で述べる。

4.1.4 旋律断片の分割方法

曲全体について網羅的に参考事例を検索するためには、様々な分割方法を用いて分割を行ない、それぞれについて検索する方がよい。理想的には可能なすべての分割を行なうのが望ましいが、ひとつの曲に対する分割の境界の引き方は音符の数のべき乗になるため計算量の観点から現実的ではない。かといってランダムに分割したのでは有効な事例をとり逃がす可能性がある。従って、ある程度一貫性のある基準に従って規則的に分割する必要がある。そこで事例に基づく演奏表情生成手法では楽曲の構成情報（楽曲構成）に従って楽曲を旋律断片に分割する。本項では楽曲構成と楽曲構成に基づく旋律断片の分割について概説する。

殆どどの楽曲は、主題、繰り返し、フレーズ、小節などの様々な大きさの部品 s の組み合わせによって構成されている。それぞれの部品は、いくつかの音符 n の集合である（式 (4.7)）。楽曲構成とは、楽曲を形作っているこれらの部品の構成のことである。なお、この楽曲構成は楽曲構造と呼ばれることもあるが、2.1.3 項で概説した認知的楽曲構造とは異なるものである。事例に基づく演奏表情生成では、この楽曲を構成する部品を旋律断片として利用する。

$$s = \{ n_i \mid n_i \text{ は部品 } s \text{ に含まれる音符} \} \quad (4.7)$$

楽曲を構成する部品は、その大きさや包含関係によって分類することができる。例えば、フレーズは小節が複数連なって構成されていることから、フレーズは小節よりもひとまわり以上大きく、小節を包含する部品である。事例に基づく演奏表情生成手法では、このような部品の大きさの違いを区別するために、旋律断片に階層という概念を導入する。大きさが概ね同じ旋律断片の集合を階層と呼ぶ(図 4.5)。分割の単位が最大の旋律断片の集合を第 1 層、 i 番目に大きい分割単位の旋律断片の集合を第 i 階層と呼ぶ。



図 4.5: 楽曲構成の境界と階層

最下層の旋律断片は、音符や休符など、楽曲を構成する最小の部品、つまり音符の集合である。最下層の一つ上の階層は音符よりひとまわり大きな部品の集合である。楽曲によって異なるが、多くの場合は一拍の大きさの部品からなる。以下、半小節、小節、フレーズ、フレーズ列、主題などのように、階層が高くなるに従って階層に含まれる旋律断片の大きさが大きくなる。最上層(第 1 層)の旋律断片は楽曲を構成する最大の部品、つまり楽曲そのものである。

第 i 層に属するある旋律断片 $s_{i,j}$ に対して、一つ上の階層(第 $i-1$ 層)に含まれていて且つ $s_{i,j}$ を包含する旋律断片を $s_{i,j}$ の親と呼ぶ(式 (4.8))。 i 層より上の階層に属していて $s_{i,j}$ を包含する旋律断片の集合を $s_{i,j}$ の祖先と呼ぶ(図 4.5, 式 (4.9))。一つ下の階層(第 $i+1$ 層)に属していて $s_{i,j}$ を構成している旋律断片の集合を $s_{i,j}$ の子、 i より下の階層に属していて $s_{i,j}$ を構成している旋律断片の集合を $s_{i,j}$ の子孫と呼ぶ(式 (4.10), 式 (4.11))。

$$\text{親}(s_{i,j}) \equiv s_{i-1,k}, \text{ where } s_{i,j} \subset s_{i-1,k} \quad (4.8)$$

$$\text{祖先}(s_{i,j}) \equiv \{s_{k,l} \mid k < i, s_{i,j} \subset s_{k,l}\} \quad (4.9)$$

$$\text{子}(s_{i,j}) \equiv \{s_{i+1,k} \mid s_{i+1,k} \subset s_{i,j}\} \quad (4.10)$$

$$\text{子孫}(s_{i,j}) \equiv \{s_{k,l} \mid k > i, s_{k,l} \subset s_{i,j}\} \quad (4.11)$$

楽曲構成の一例として一般的なポピュラー音楽などの楽曲に見られる楽曲構成を図 4.6 に示す。この楽曲は、「一番」「二番」という主要な部品を「前奏」「間奏」「後奏」という繋ぎで挟み込む形で曲全体が構成されている。一番と二番は同じような構成になっており、それぞれ「A メロ」「B メロ」「サビ」という 3 つの部品が順番にならんでいる。さらに、「サビ」の部分に注目すると同じようなフレーズが 2 回繰り返されている。この楽曲では、前奏、一番、間奏、二番、後奏、A メロ、B メロ、サビ、サビのフレーズなどの音楽的に意味のある部品から構成されている。また、曲全体も音楽的に意味を持ったまとまりであることから、楽曲を構成する部品の一種と考えることができる。

この楽曲を構成する部品を大きさや包含関係に基づいて順に並べると次の様になる。

1. 曲全体
2. 前奏、一番、間奏、二番、後奏
3. A メロ、B メロ、サビ

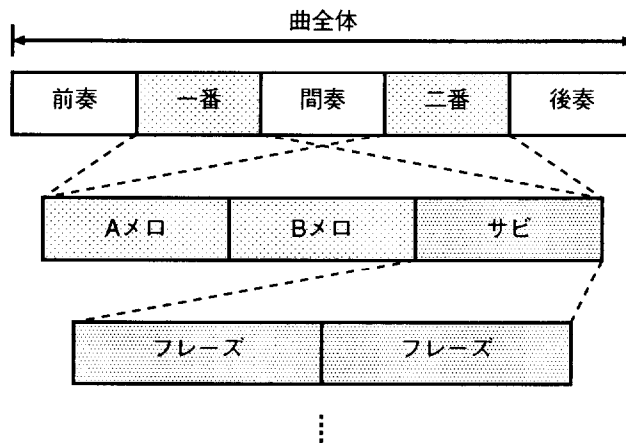


図 4.6: 楽曲構成の例

4. サビのフレーズ

実際には、前奏、間奏、後奏や A メロ、B メロなども様々な部品から構成されており、省略されている部分を補うと、末端が音符となるような木構造となる。これが、階層構造の各階層に含まれる部品である。これがそのまま旋律断片となる。第 1 層に属する旋律断片は曲全体である。第 2 層には、前奏、間奏、二番、後奏が属する。一番の親は曲全体で、一番の子は、A メロ、B メロ、サビの 3 つである。また、サビの祖先は、一番（もしくは二番）と曲全体である。

楽曲構成情報は、楽曲の分割方法として楽譜上などに間接的に記載されているため、比較的容易に取得することが可能である。また、分割の手法に一貫性があるためランダムな分割などに比べ参考事例の検索の際により類似した旋律断片が見つかる可能性が高い。さらに、一般に楽曲構成は拍節構造を内包していることが多いため、楽曲構成に基づいて区切られた旋律断片どうしは拍節的にも類似した特徴を持ちやすく、類似した参考事例の検索に適している。

一方で、4.2 節で後述する演奏表情の記述法（演奏表情の大局比率）を利用するためには、楽曲に対し 4.2.2 項に示すような包含関係を持った境界情報を定義する必要がある。楽曲構成は、演奏表情の大局比率が用いる境界情報の必要条件を比較的容易に満たすことが可能であるため、演奏表情の大局比率の境界情報としても好都合である。なお、大局比率の境界情報の条件については 4.2.2 項 で述べる。

4.1.5 演奏データの有効利用を考慮した参考事例検索

ここでは、本節で提案した事例の有効利用手法を、事例に基づく演奏表情生成手法においてがどのように利用するかについて解説する。

本節で述べた参考事例の有効利用に関する技術の要点は以下の 2 つである。

- 演奏対象曲を楽曲構成情報に基づいて様々な長さの旋律断片に分割し、旋律断片ごとに類似旋律を検索する
- 参考事例検索時の旋律断片の類似性の評価を旋律断片の全体的な特徴の類似性のみから判定する

事例に基づく演奏表情生成手法においてこれらが直接関係するのは、参考事例の検索処理である。しかしながら、本節で述べた技術は処理対象となる旋律断片や演奏表情に対して制約を与えるため、結果的に

検索以降の処理でもこの制約を考慮する必要がある。以降では、事例に基づく演奏表情生成を構成する各々の処理ごとに、事例を有効利用する手法や、それに伴って配慮すべき点について述べる。

参考事例の検索 参考事例の検索は、参考事例の有効利用に関する技術に最も関連した処理である。事例の参考事例の検索は以下のように行なわれる。

- 演奏データ集を旋律断片に分割
- 演奏対象曲を旋律断片に分割
- 演奏対象曲の全ての旋律断片に対して、演奏データ集から参考事例を検索

まず最初に、演奏データ集の全ての演奏データについて、4.1.4 項で述べた楽曲構成情報に基づいて境界の決定及び分割を行なう。この時点で、演奏データ集は楽曲に対する演奏データの集合ではなく、旋律断片の演奏データの集合になる。また、演奏対象曲についても同様に楽曲構成情報に基づいて分割を行なう。そして、演奏対象曲を分割した結果得られる全ての旋律断片に対して、演奏データの旋律断片集合の中から旋律が類似した旋律断片の演奏データを検索し、その結果を旋律断片ごとの参考事例集としてこれ以降の処理で利用する。

重要度の評価 参考事例を検索する際には、4.1.3 項で述べたように、旋律断片の全体的な特徴から旋律の類似性を判断したが、重要度の評価の際には、旋律の全体的な特徴だけではなく、子細な部分も考慮して旋律の類似性を評価した方がよい。

参考事例の検索で、旋律断片の全体的な特徴のみを利用するのは、検索によって得られる参考事例集の大きさを十分なものにするためである。一方で、ここで計算する重要度は、対象曲の演奏表情生成を行なう上でそれぞれの参考事例がどの程度有用であるかの指標である。この有用性の判断をよりの確に行なうためには、可能な限り多くの情報を導入し、旋律の類似性を総合的に判断する必要がある。従って、子細な部分の類似性も考慮した方がよい。

このことは図 4.7 のような場合を考えるとわかりやすい。対象曲のある旋律断片に対する参考事例集の中に、全体的な特徴が全く同じ参考事例が二つあり、片方は子細な部分が対象曲の旋律断片にあまり類似していないが、もう一方は子細な部分も類似していたとする。この場合、子細な部分まで類似している参考事例の方が、対象曲の演奏表情生成に有用であることは明らかである。

しかし、全体の特徴があまり似ていないにもかかわらず、子細な部分が類似しているため、重要度が他のある程度全体的な特徴が似ている事例より高くなってしまいうという事態も好ましくはない。参考事例から取り出されるのはあくまで全体的な演奏表情のみであるため、基本は全体的な類似性を最優先にして比較し、それで優劣がつきにくい場合にそれを補う目的で子細な部分の類似性を考慮すべきである。

演奏表情の取り扱い 4.1.3 項で述べたように、個々の参考事例からは全体的な演奏表情のバランスのみを抽出して利用する。このためには、演奏表情の分析の際には、まず、旋律断片内の演奏表情の変化を、全体的な変化と部分的な変化に分離する必要がある。そこで、事例に基づく演奏表情生成では、このような計算が可能な演奏表情の表現手法として演奏表情の大局比率(4.2 節 参照)を導入する。従って、演奏表情の取り扱いを行なっている部分では、この大局比率表現の上で必要な計算を行なう必要がある。

事例に基づく演奏表情生成手法で具体的にこれらが関係する処理は、以下の三つである。

- 参考事例の演奏表情の分析
- 演奏対象曲の演奏表情の合成
- 演奏対象曲への演奏表情の適用

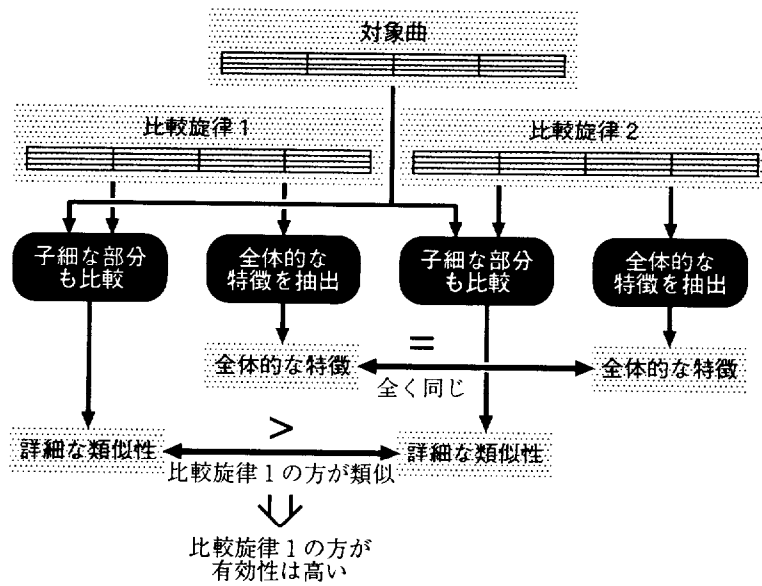


図 4.7: 全体的な特徴のみから参考事例の重要度を評価することの問題

これらの処理で演奏表情の大局比率をどのように利用するかについては、4.2.3 項で説明する。

4.2 演奏表情の記述と計算

4.1.3 項で述べたように、大きな旋律断片に対する事例の不足を回避するため、本手法では旋律断片ごとの参考事例の検索の際に旋律断片の全体的な特徴のみに注目する。しかし、全体的な特徴が類似しているだけの事例は、子細な部分の類似性が保証されていないため、子細な部分に対する演奏表情（局所的な演奏表情）の参考事例として利用するのは適切ではない。そこで、事例に基づく演奏表情生成手法では、参考事例中の演奏表情から、局所的な演奏表情を除去し、全体的な演奏表情の変化のみを取り出す（図 4.8）。

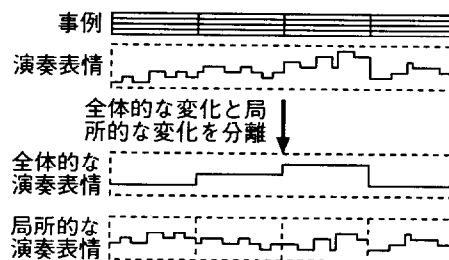


図 4.8: 全体的な演奏表情と局所的な演奏表情の分離

本節では、参考事例の演奏表情を局所的な演奏表情と全体的な演奏表情に分離することが可能な演奏表情の表現法として、演奏表情の大局比率を提案する。また、演奏表情の大局比率を利用した演奏表情の計算についても述べる。

4.2.1 演奏表情の表現

演奏表情生成に限らず、計算機で演奏表情を取り扱うためには、テンポや音の強さなどの演奏表情の絶対的な数値（演奏表情値）を何らかの形式でデータとして記述する必要がある。

最も一般的な演奏表情値の記述法は、一音ごとにその絶対的な数値を記録するという手法である。例えば、テンポであれば楽譜上の一音ごとにそのテンポ値を記すというものである。テンポ値は楽曲全体の平均演奏速度を表すことに使われることが多いが、一小節や一音といったより短い範囲での値を計算することも可能である。テンポ値は、式 (4.12) に示すように、基準音符の拍数を演奏時間で割ったものである。基準音符とは、テンポ表示に表記されるテンポ計算の基準となる音符である。通常は、拍子記号の母数をそのまま利用する。基準音符の拍数はテンポを求める楽曲に先頭から基準音符を並べていくと、全部でいくつ入るかを数えたものである。例えば、四分の六拍子で全部で 60 小節からなる楽曲の場合、基準音符が四分音符で基準音符の拍数が 360 となる。この楽曲の演奏時間が 4 分だった場合、テンポは $360/4 = 90$ である。

$$\text{テンポ} = \frac{\text{基準音符の拍数}}{\text{演奏時間 (分)}} \quad (4.12)$$

式 (4.12) は、楽曲全体でなくても適用可能である。楽曲中のある区間に対して、その区間の基準音符の拍数と演奏時間を測定することができれば、その区間のテンポ値を計算することができる。このような、楽曲の一部区間でのテンポ値を局所テンポと呼ぶ。一音ごとのテンポ値による表記は、テンポの計算区間を楽譜上の音符ひとつに設定し、全音符について計算したものである。

音の強さについても同様の表記が可能である。音の強さの場合、打鍵時の圧力や鍵の移動速度を測定して記録するが、このような情報はもともと一音単位で測定せざるを得ないことから、必然的に一音ごとの演奏表情値が得られる。

さて、本節の冒頭でも述べたように、事例に基づく演奏表情生成手法で用いる参考事例は、子細な部分での類似性が保証されていないため、子細な部分に対する演奏表情を取得するには適さない。そこで、参考事例中の演奏表情から、子細な部分に対する演奏表情と全体的な演奏表情の変化を分離し、全体的な演奏表情のみを取り出す必要がある。一方で、一般的に使われている一音単位の演奏表情値は、楽曲中の最も子細な部分に対する演奏表情の情報であるから、これをそのまま利用することはできない。

一音ごとの演奏表情値から演奏表情の全体的な変化を取り出す手法として、演奏表情値の変化量を曲線近似する方法が知られている。一般的な手法は、小節やフレーズなどの長さを決め、その区間内の演奏表情の変化を、最小自乗法などを用いて演奏表情の変化を 2 次や 3 次の近似曲線に近似するというものである。近似によって演奏表情の微小な変化が除去され、全体的な区間の長さを様々に変更すれば、様々な長さの旋律に対する演奏表情の全体的な変化を、近似曲線のパラメータという形で取り出すことができる。

同様の手法として、典型的な変動パターンをいくつか用意しその中の近いものを割り当てるという手法 [64] がある。この手法では、まず、典型的な演奏表情の変動を表す曲線「変動パターン」を何種類か用意する。変動パターンは、「登り」、「下り」、「丘」、「盆地」など、図 4.9 に示すような、2 ないしは 3 次程度の複雑さの曲線である。この変動パターン曲線の中から実際の演奏表情の変化に近い曲線を見つけだし、その区間の演奏表情の全体的な演奏表情の変化として割り当てていく。小節やフレーズなど、様々な長さの区間の演奏表情に対して、それぞれ近い曲線を見つけだし、その組合せで楽曲全体の演奏表情の傾向を表現することもできる。近似曲線を用いる手法では関数のパラメータという抽象度の高い情報に帰着しているのに対し、この手法では曲線の概形という、人間の感覚に近い特徴情報を取り出すことができる。

これらの、曲線に近似する手法は、全体的な演奏表情と子細な部分の演奏表情の分離が容易であるという点では、事例に基づく演奏表情生成手法にも適している。しかしながら、曲線で近似するとともとの演奏中に見られる演奏表情の情報を完全に正確に表現できるわけではない。また、演奏表情値を連続的な値として曲線に近似すること自体の問題も指摘されている [11]。

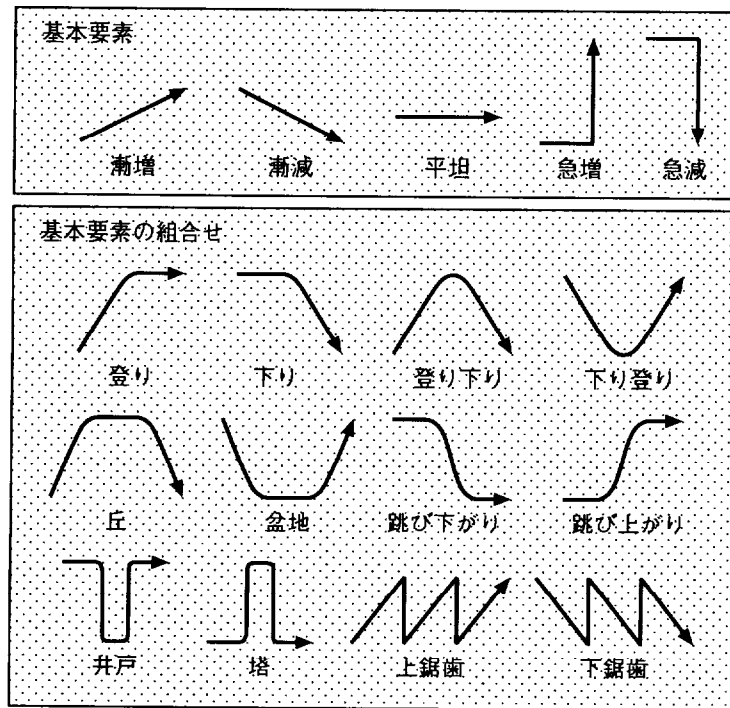


図 4.9: 演奏表情の変動パターン曲線

4.2.2 演奏表情の大局比率

演奏表情の大局比率は、演奏表情値を一音ごとの数値ではなく、再帰的且つ階層的な比率で表現したものである。図 4.10 に示すように、境界情報を利用して分割された楽曲の区間内での局所的な演奏表情値と、その区間を含んでいて且つ一回り大きな区間での平均値との比率で表される。

演奏表情の大局比率を利用するためには楽曲の境界情報を与える必要がある。これは、従来の曲線近似法などで用いられている近似を行なう区間とほぼ同じ役割の情報で、大局比率を求める際に計算範囲の基準として利用する。境界情報は、以下の条件を満たさなければならない(図 4.11)。

条件 1 分割単位の大きさの異なる何種類かの境界集合が必要

条件 2 任意の境界集合同士に包含関係がある

条件 1 は、境界情報 B は分割単位が異なる複数の境界集合 B_i の集合であるということである(式 (4.13))。境界集合とは、対象曲を複数の断片に重複なく分割する境界の拍位置 $b_{i,j}$ の集合である(式 (4.14), (4.15))。 l は楽曲の総拍数、 m_i は境界の総数を表す。任意の境界集合には楽曲の先頭と末尾を必ず含むものとする(式 (4.16), (4.17))。境界集合の分割単位 u_i とは境界集合によって区切られた個々の区間の長さの平均値を表す(式 (4.18))。例えば、分割単位が一小節の境界集合は、楽曲中の全ての小節線の集合である。個々の区間に重複がないことから、境界の総数 m_i と分割単位 u_i との積は楽曲の長さ l に一致する。この境界集合を複数個 (n 個) 集めた集合が境界情報である。図 4.11 に例示した境界情報は 4 種類の境界集合から構成されている。

$$B = \{B_i \mid 1 \leq i \leq n\} \quad (4.13)$$

$$B_i = \{b_{i,j} \mid 0 \leq b_{i,j} \leq l, 0 \leq j \leq m_i\} \quad (4.14)$$

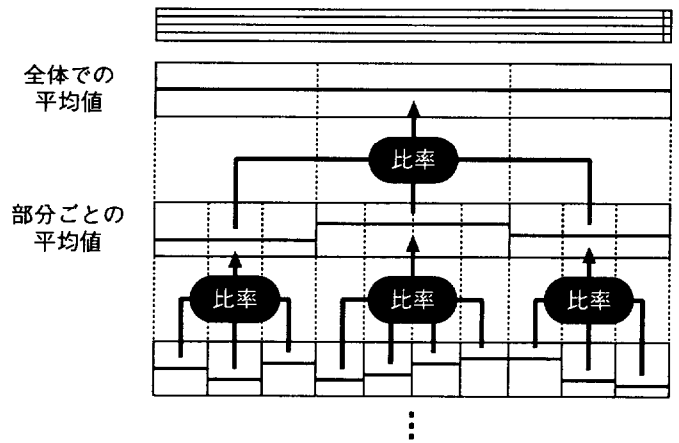


図 4.10: 演奏表情の大局比率の概念

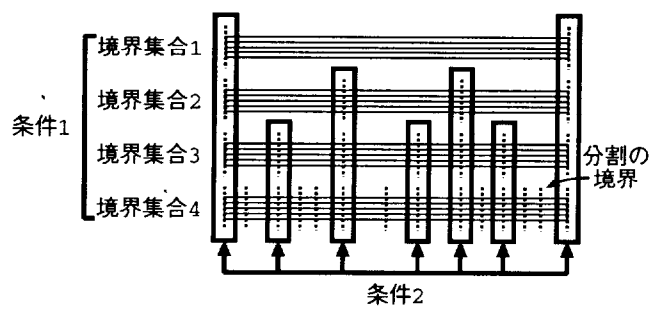


図 4.11: 演奏表情の大局比率で利用する境界情報の条件

$$b_{i,j} < b_{i,k} \text{ if } j < k \quad (4.15)$$

$$b_{i,0} = 0 \quad (4.16)$$

$$b_{i,m_i} = l \quad (4.17)$$

$$u_i = \frac{\sum_{j=1}^{m_i} b_{i,j} - b_{i,j-1}}{m_i} \quad (4.18)$$

条件 2 は、全ての境界集合の間に包含関係（式 (4.19)）があるということである。従って、包含関係を利用して境界集合の間に大小関係を定義することが可能である。以降では、便宜のため境界集合 B_i は小さい順に並んでいるものとする。

$$B_i \subset B_j \equiv \forall k \ b_{i,k} \in B_i \rightarrow b_{i,k} \in B_j \quad (4.19)$$

大局比率を演奏表情の記述法として用いるだけであれば、境界情報が音楽的に意味のあるものである必要はないが、大局比率から音楽的な特徴や傾向を抽出する場合は、楽曲構成などのような音楽的に意味のある境界情報を与えた方がよい。

事例に基づく演奏表情生成手法では、演奏表情の大局比率を全体的な演奏表情と子細な部分の演奏表情との分離という音楽的な意味を持った情報として利用しているため、境界情報として 4.1.4 項で述べた楽曲構成に基づく旋律分割を利用している。この場合、楽曲構成の各階層がそれぞれ境界集合に該当する。殆んどどの楽曲は、上記の境界条件を満たす楽曲構成情報を作成することが可能である。以降の説明では、境界情報によって区切られた区間を旋律断片と呼ぶ。また、旋律断片の関係も 4.1.4 項において示した「親」「子」などの表記を用いる。

次に、大局比率の計算法を解説する。演奏表情の大局比率 R は、境界集合 B_i によって区切られた区間 $(b_{i,j}, b_{i,j+1})$ の局所的な演奏表情値と、それよりひとつ小さな境界集合 B_{i-1} で区切られ $(b_{i-1,k}, b_{i-1,k+1})$ を包含する区間 $(b_{i-1,k}, b_{i-1,k+1})$ の局所的な演奏表情値との比率である（図 4.12）。演奏表情 exp の大局比率 R_{exp} を式 (4.20) に示す。 $exp(b_1, b_2)$ は b_x 拍から b_y 拍までの区間の局所的な演奏表情値を表す。

$$R_{exp}(b_{i,j}, b_{i,j+1}) = \frac{exp(b_{i,j}, b_{i,j+1})}{exp(b_{i-1,k}, b_{i-1,k+1})} \quad (4.20)$$

$$0 \leq j < m_i, \quad 0 \leq k < m_{i-1}$$

$$b_{i-1,k} \leq b_{i,j}, \quad b_{i,j+1} \leq b_{i-1,k+1}$$

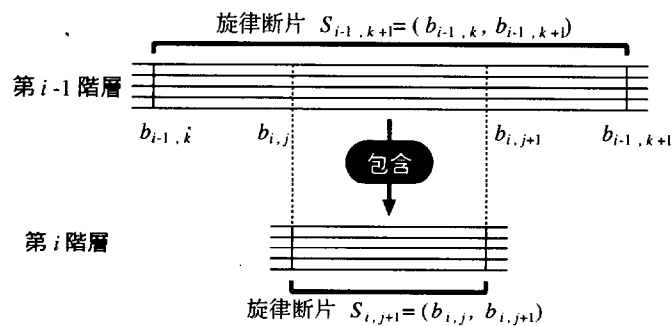


図 4.12: 演奏表情の大局比率の計算

これは、4.1.4 項で述べた楽曲構成に当てはめれば、第 i 階層の旋律断片 $s_{i,j}$ での局所的な演奏表情値とその親の断片 $s_{i-1,k}$ での演奏表情値の平均値との比率ということになる（式 (4.21)）。

$$R_{exp}(s_{i,j}) = \frac{exp(s_{i,j})}{exp(s_{i-1,k})} \quad (4.21)$$

$$s_{i-1,k} = \text{親}(s_{i,j})$$

テンポのように旋律断片の演奏表情値が直接取得できる場合は、上記の $exp(s_{i,j})$ が旋律断片の演奏表情値そのものになる (式 (4.22))。

$$R_{tempo}(s_{i,j}) = \frac{tempo(s_{i,j})}{tempo(s_{i-1,k})} \quad (4.22)$$

音の強さなどのように一音ごとの演奏表情値が得られる演奏表情の場合は、各音の演奏表情値を平均化した上で式 (4.21) に適用する。旋律断片 $s_{i,j}$ の音の強さに関する演奏表情の大局比率 $R_{dynamics}(s_{i,j})$ を式 (4.23) に示す。 n は $s_{i,j}$ に含まれる音符、 $|s_{i,j}|$ は $s_{i,j}$ に含まれる音符の総数、 $dynamics(n)$ は音符 n の音の強さを表す。

$$R_{dynamics}(s_{i,j}) = \frac{\sum_{n \in s_{i,j}} dynamics(n)}{|s_{i,j}|} \cdot \frac{|s_{i-1,k}|}{\sum_{n \in s_{i-1,k}} dynamics(n)} \quad (4.23)$$

この計算を、楽曲に含める全階層の全ての旋律断片に対して行なった結果の集合が演奏表情の大局比率である。

演奏表情の大局比率を用いて演奏表情を記録する場合は、境界情報によって分割された全ての区間について大局比率の数値を計算し、これに曲全体での平均値を付記することで、楽曲中の演奏表情の変化を表記する (図 4.13)。

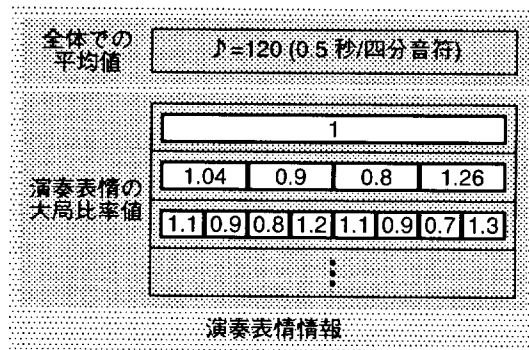


図 4.13: 演奏表情の大局比率を用いた演奏表情の記録

演奏表情の大局比率の計算例として、図 4.14 に 4 小節からなるフレーズの小節ごとのテンポについての演奏表情の大局比率の計算例を示す。このフレーズは、フレーズ全体が一つの断片であり、また、各小節がフレーズ全体より一回り小さい断片である。このフレーズは全体の平均テンポが 120 で、それぞれの小節の平均のテンポが順番に 115, 133, 150, 95 で演奏されたとする。

テンポは演奏時間の逆数であるため、平均などの計算を行なうと実際の値と一致なくなってしまう。上記の例でも 115, 133, 150, 95 の単純平均は 123.25 となり 120 とは一致しない (表 4.1)。よって、テンポに限っては演奏表情の大局比率の計算にテンポの逆数である「一拍の演奏時間」を利用する。一拍の演奏時間の計算式を式 (4.24) に示す。一拍の演奏時間の単位は s/beat (秒/拍) である。

$$\text{一拍の演奏時間} = \frac{60}{tempo(s)} \quad (4.24)$$

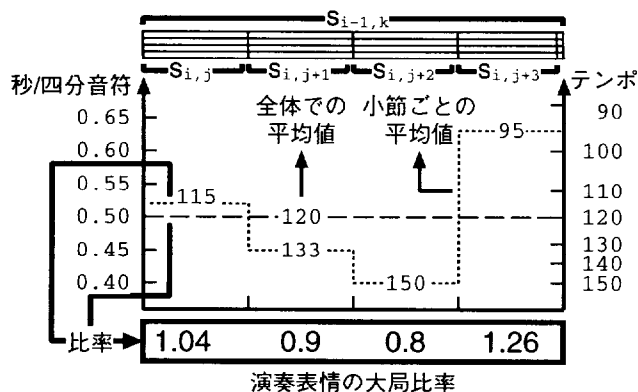


図 4.14: テンポに関する演奏表情の大局比率の計算例

図 4.14 の例の場合、フレーズ全体の一拍の演奏時間が 0.5 s/beat, 各小節の一拍の演奏時間はそれぞれ 0.52 s/beat, 0.45 s/beat, 0.4 s/beat, 0.63 s/beat となる。

各小節の演奏表情の大局比率は、各小節の演奏表情値とその親であるフレーズ全体の演奏表情値の平均との比率である。よって、この例の各小節のテンポに関する演奏表情の大局比率は、順に 1.04, 0.9, 0.8, 1.26 である。

	テンポ	一拍の演奏時間	演奏表情の大局比率
$s_{i-1,k}$	120	0.50	—
$s_{i,j}$	115	0.52	1.04
$s_{i,j+1}$	133	0.45	0.90
$s_{i,j+2}$	150	0.40	0.80
$s_{i,j+3}$	95	0.63	1.26
平均	123.25	0.50	1.00

表 4.1: テンポに関する演奏表情の大局比率の計算例

実際の楽曲では、小節の子孫である個々の拍や音符の演奏表情の大局比率や、フレーズの祖先の旋律断片についても、再帰的に演奏表情の大局比率を計算する。

大局比率は親の演奏表情値の平均に対する子の平均値の比率であるが、多くの場合、親の平均値はその全ての子の演奏表情値の平均値になっているため、全ての子の大局比率の平均は 1 になる。表 4.1 に示す計算例でも、旋律断片 $s_{i,j} \cdots s_{i,j+3}$ の大局比率の平均が 1 になっている。

演奏表情の大局比率の計算では近似を行っていないため、大局比率で記述された演奏表情情報から一音ごとの絶対的な演奏表情値を復元することが可能である。演奏表情の大局比率から一音ごとの絶対的な演奏表情値への逆変換計算について説明する。

先に述べたように、ある旋律断片 $s_{i,j}$ の演奏表情 exp の大局比率 $R_{exp}(s_{i,j})$ は、親となる旋律断片の演奏表情値に対して、注目している旋律断片 $s_{i,j}$ の演奏表情値がどの程度の比率であるかを表したものである。従って、親となる旋律断片 $s_{i,j}$ の局所的な演奏表情値が算出できれば、式 (4.25) に示す式により注目する旋律断片 $s_{i,j}$ の局所的な演奏表情値 $exp(s_{i,j})$ も計算することができる。

$$\exp(s_{i,j}) = R_{exp}(s_{i,j}) \cdot \exp(\text{親}(s_{i,j})) \quad (4.25)$$

親の旋律断片の演奏表情値は、親の親の旋律断片の演奏表情値がわかれば式(4.25)式により計算することができる(式(4.26))。

$$\exp(\text{親}(s_{i,j})) = R_{exp}(s_{i,j}) \cdot \exp(\text{親}(\text{親}(s_{i,j}))) \quad (4.26)$$

従って、この計算を再帰的に繰り返せば最終的に最大の旋律断片である楽曲全体にたどり着く。最終的に楽曲全体での演奏表情の平均値が必要となるが、本項の演奏表情の大局比率の計算の説明の際に述べたように、演奏表情の大局比率では、各旋律断片の比率値と、楽曲全体での演奏表情の平均値の両方を記録することから、この値は既知である。従って、曲全体での演奏表情の平均値 $\exp(s_{1,1})$ に $s_{i,j}$ の全ての祖先の演奏表情の大局比率を掛け合わせたものが、 $s_{i,j}$ の演奏表情値となる(式(4.27))。

$$\exp(s_{i,j}) = \exp(s_{1,1}) \cdot \prod_{s \in A} R_{exp}(s) \quad (4.27)$$

$$A = \text{祖先}(s_{i,j})$$

楽曲中の個々の音符は最下層の旋律断片であることから、音ごとの絶対的な演奏表情値もこの式により復元することが可能である。

演奏表情の復元の例として、図 4.15 に 4 小節の長さからなる楽曲のテンポ値の復元を示す。この楽曲は、それぞれ楽曲全体、小節、半小節の長さからなる 3 階層の旋律断片で構成されている。この例では、半小節ごとの演奏表情値の復元を行なう。各旋律断片の演奏表情の大局比率の値は、図中に示す通りである。

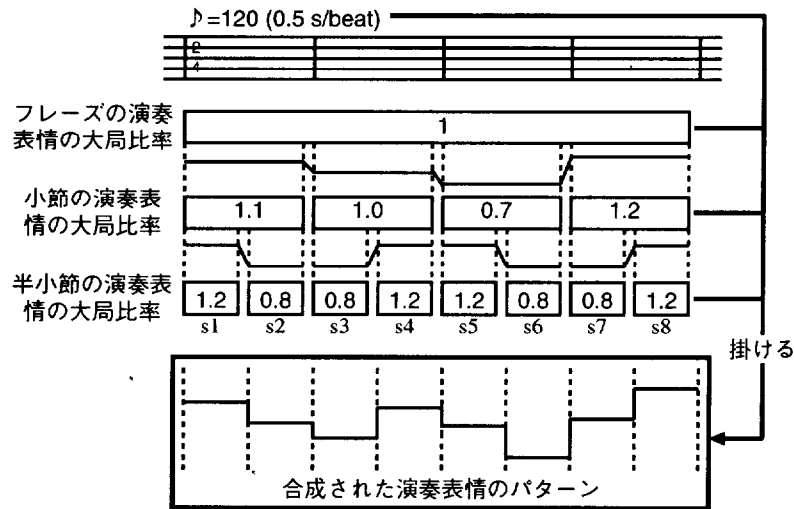


図 4.15: 演奏表情の大局比率から絶対的な演奏表情値への復元の例

半小節の演奏表情の大局比率は小節の演奏表情値と半小節の演奏表情値の比率である。また、小節の演奏表情の大局比率はフレーズの演奏表情値と小節の演奏表情値の比率である。従って、半小節の演奏表情の大局比率と小節の演奏表情の大局比率を掛け合わせることで、フレーズの演奏表情値と半小節の演奏表情値の比率が求まる。例えば、一番最初の半小節の場合、楽曲全体のテンポに対する比率は、半小節自身の演奏表情の大局比率 1.2 と親となる小節の大局比率 1.1、1 の積 1.32 である。これにフレーズ全体の演奏表情値を掛けると、半小節の演奏表情値が得られる。大局比率の計算の際に述べたように、テンポ

計算の場合はテンポ値ではなく一拍の演奏時間式 (4.24) を用いる。この例の場合、楽曲全体の一拍の演奏時間が 0.5 なので、一番最初の半小節のテンポは $1.32 \times 0.5 = 0.66$ (s/beat) = 90.9 となる。個々の半小節ごとの演奏表情値の復元結果を表 4.2 に示す。

半小節	大局比率の積	一拍の演奏時間	テンポ
s1	1.32	0.66	90.9
s2	0.88	0.44	136.4
s3	0.8	0.4	150.0
s4	1.2	0.6	100.0
s5	0.84	0.42	142.9
s6	0.56	0.28	214.3
s7	0.96	0.48	125.0
s8	1.44	0.72	83.3

表 4.2: 大局比率の復元の結果

演奏表情の大局比率を用いた表記は以下の特徴を持つ。

- 注目している音符や区間の演奏表情値の相対的な比率を表す
演奏表情の大局比率は演奏表情値の比率を表していることから、相対的な演奏表情の関係を知らることが可能である。
- 全体的な演奏表情と局所的な演奏表情を分離することが可能
全体的な演奏表情の変化は長い区間の大局比率値に表され、局所的な演奏表情はその影響する範囲に応じた短い区間の大局比率値に現れる。従って、演奏表情の大局比率では全体的な演奏表情の変化と子細な部分で起こっている局所的な変化を別個の数値へと分離することが可能である。このことから、演奏表情の大局比率は 4.1.3 項 で述べた演奏表情分析に求められる要求を満たすものである。
- 元の各音の演奏表情値を完全に復元することが可能
一音ごとの演奏表情値から演奏表情への大局比率への変換の過程では近似を行っていないため、この変換は可逆変換である。従って、演奏表情の大局比率からもととの各音の演奏表情値を完全に復元することが可能である。

4.2.3 事例に基づく演奏表情生成手法での演奏表情の大局比率

事例に基づく演奏表情生成手法で、演奏表情の大局比率が関係するのは以下の三つの処理である。

- 参考事例の演奏表情の分析
- 演奏対象曲の演奏表情の合成
- 演奏対象曲への演奏表情の適用

参考事例の演奏表情の分析 参考事例の分析では、参考事例集に含まれる個々の旋律断片の演奏表情の大局比率の計算を行なう。この計算は 4.2.2 項 で述べた通りである。テンポや音の強さなど演奏表情値ごとに上記の計算を行ない、その組合せを旋律断片の演奏表情の大局比率として記録する。

演奏対象曲の演奏表情の合成 合成手続きでは、個々の参考事例の旋律断片の演奏表情の大局比率を合成し、演奏対象曲の旋律断片の大局比率値を作成する。演奏表情の合成の時点で、対象曲に対する全ての参考事例について、演奏表情の大局比率値と事例の重要度が得られている。演奏対象曲の旋律断片の演奏表情の大局比率値は、参考事例の演奏表情の大局比率値の重要度による加重平均である。大局比率は比率であるため、そのまま平均などの計算を行なうことができる。

対象曲の旋律断片 $s_{i,j}$ の演奏表情 exp の大局比率値 $R_{exp}(s_{i,j})$ の計算式を式 (4.28) と図 4.16 に示す。 E は重要度の評価や演奏表情の分析を行なった後の参考事例集である。 $E_{i,j}$ は E 中の $s_{i,j}$ に類似した旋律の事例の集合である。 e は個々の事例を表す。 w_e 、 $R_{exp,e}$ はそれぞれ事例 e の重要度と大局比率の数値である。

$$R_{exp}(s_{i,j}) = \frac{\sum_{e \in E_{i,j}} w_e \cdot R_{exp,e}}{\sum_{e \in E_{i,j}} w_e} \quad (4.28)$$

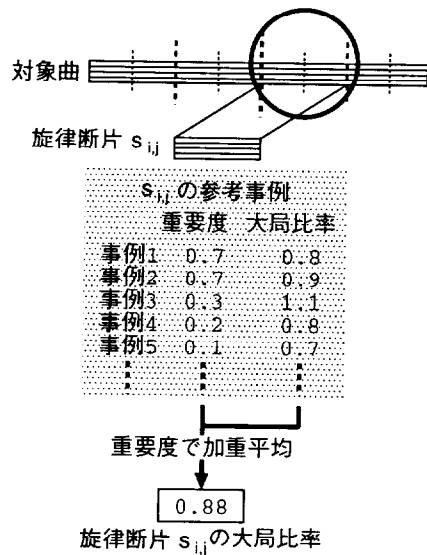


図 4.16: 演奏表情の大局比率の合成

演奏対象曲への適用 演奏対象曲の適用処理では、合成処理によって作成された演奏対象曲の個々の旋律断片の大局比率 $R_{exp}(s_{i,j})$ から、演奏対象曲の一音ごとの演奏表情値を復元する。ここでは、旋律断片ごとに大局比率で表記された演奏表情を、音符ごとの絶対的な演奏表情値に変換する操作を行なう。4.2.2 項で述べた通り、楽曲の全ての旋律断片の演奏表情の大局比率が既知であれば、楽曲の任意の旋律断片について、絶対的な演奏表情値を式 (4.27) に示す式で復元することが可能である。音符も旋律断片の一つであることから、音符についても同様である。

演奏対象曲の i 番目の音符 $s_{n,i}$ の演奏表情値 $exp(s_{n,i})$ の計算を式 (4.29) に示す。 n は楽曲構成の階層の総数である。音符は最下層の階層の構成要素であるため、 i 番目の音符は再下層の i 番目の構成要素 $s_{n,i}$ で表される。演奏表情の復元計算には曲全体での演奏表情の平均値 $exp(s_{1,1})$ が必要があるが、これは曲全体に対する参考事例から取得する。テンポや音の強さなどの演奏表情ごとに、この一連の計算を行ない、一音ごとの演奏表情値を算出する。

$$\exp(s_{n,i}) = \exp(s_{1,1}) \cdot \prod_{s \in \text{祖先}(s_{n,i})} R_{\exp}(s) \quad (4.29)$$

4.3 旋律の類似性の評価

4.1.3 項で述べたように、大きな旋律断片の事例を検索する際に、旋律内の子細な部分の類似性は考慮せずに、全体的な特徴の類似性のみ注目する。本節ではこのような全体的な特徴に基づく旋律の類似性の評価手法について述べる。

4.3.1 旋律の類似性の評価

旋律の類似性の評価に関する研究は、音楽データベースの検索の研究分野を中心に様々な提案がなされている。

古典的な旋律比較の手法では、旋律の先頭から漸進的に比較を行なう。最も単純な手法としては、比較する二旋律の音高の推移の差を積分し、その値の大小から旋律の類似性を評価する手法 [79] が知られている (図 4.17)。この手法では、各々の旋律の長さが同じでなければならない。また、音高の推移の比較はできるが、拍数や音符数、音長の相違などの情報はそのままでは考慮されない。さらに、休符の扱いや音符ごとの拍節上の重要度の差を考慮するのも困難である。

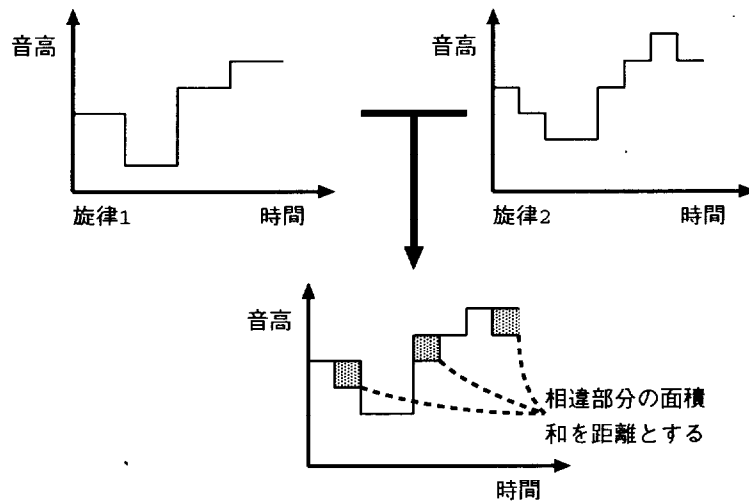


図 4.17: 音高の推移の差分を利用した旋律比較

同様の手法として、一拍や半拍ごとに音符の音高の近さを比較し、その結果を拍の重要度に応じて重み付けし、合成するという手法がある。この手法の場合、拍節上の重要度の差などを取り扱うことができる点で先の手法に比べ優れているが、音符の平均拍長が異なる場合には適切な比較を行なうのが難しい。

その他に、比較する2つの旋律を小節や拍などの一定の短い区間に分割して、その区間ごとの特徴量を計算し、この特徴量を先頭から漸進的に比較するという手法がある [49, 57] (図 4.18)。この手法では、特徴量を適切に選択すれば、休符や拍節上の重要度を考慮することが可能である。また、個々の区間が短いため区間ごとの特徴情報があまり複雑にならない。従って、区間の特徴量を用いてインデックスを作成することが可能である。インデキシングは、旋律データベースの検索の効率化に有効であるほか、同

じ楽曲の中からの繰り返し部分の発見などへの応用も可能である。一方で、漸進的な比較を行なう必要があるため、類似性の評価法として利用するためには、旋律の長さが同じでなければならない。また、基本単位より短い旋律の比較を行なうことができない。

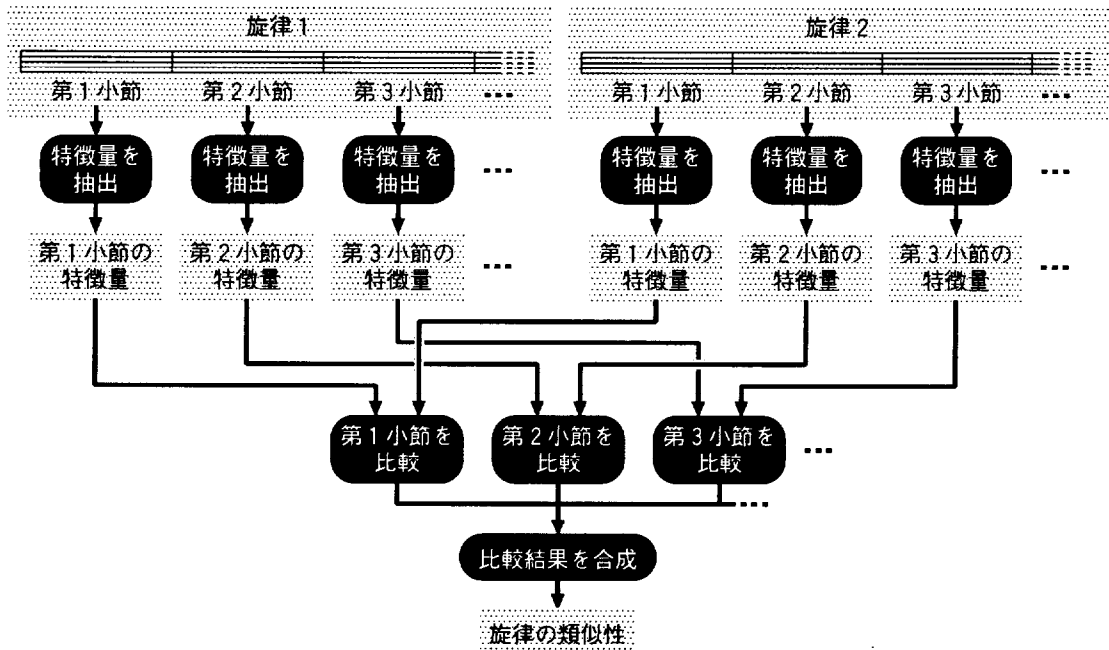


図 4.18: 小節や拍などの特徴情報を利用した比較

この他にも、文字列比較を応用した手法や動的計画法を用いた手法が提案されている [6]。もともと、旋律の類似性の評価研究は楽曲検索を目的とした研究が多い。そのためか、事例に基づく演奏表情生成手法で必要となる、全体的な類似性に重点を置いた類似性の評価方法はあまり提案されていない。

4.3.2 全体的な特徴による旋律の類似性評価

事例に基づく演奏表情生成手法では、様々な長さの旋律断片に対して類似した旋律を検索するので、可変長の入力に対応できる類似性の評価手法が必要となる。また、4.1.3 項で述べたように、従来の先頭から漸進的に比較する方法や固定長に旋律を区切って比較する手法は、旋律が長くなるに従って旋律が類似する確率が低下してしまうため、長さが長い旋律断片に対する事例の検索には有効ではない。

そこで、事例に基づく演奏表情生成手法では、旋律断片が持つ全体的な特徴から旋律断片の類似性を判定する。ここでは、旋律の全体的特徴のみに考慮した旋律類似性の評価について述べる。

ここで用いる旋律の比較手法は、旋律断片の特徴情報を抽出し、その特徴を比較することで旋律断片同士の類似性の判定を行なう。以下にその手順を示す (図 4.19)。

1. 比較する旋律の特徴量を抽出
2. 特徴量を比較して特徴量ごとの類似性を評価
3. それぞれの特徴量の類似性から、総合的な類似性を評価

まず最初に、旋律の特徴を表現するための特徴量をいくつか用意する。この特徴量は、いずれも旋律断片の長さによらずパラメータの総数が同じになるようなもので、数値の分布も長さの影響を受けにくいも

のである必要がある。特徴量を抽出したら、それぞれの旋律断片の特徴量同士を比較し、類似性の判定を行なう。ここでは、特徴量ごとにその類似性を評価する関数が必要である。特徴量は複数の数の組合せでも構わないが、特徴量の類似性評価関数の出力はスカラーでなければならない。また、特徴量ごとの類似性評価関数の出力の分布ができるだけ似ている方がよい。そして、それぞれの特徴量ごとの類似性の評価結果を総合し、最終的な旋律の類似性として出力する。この際に、それぞれの特徴量の類似性の正規化や重み付けが必要である。

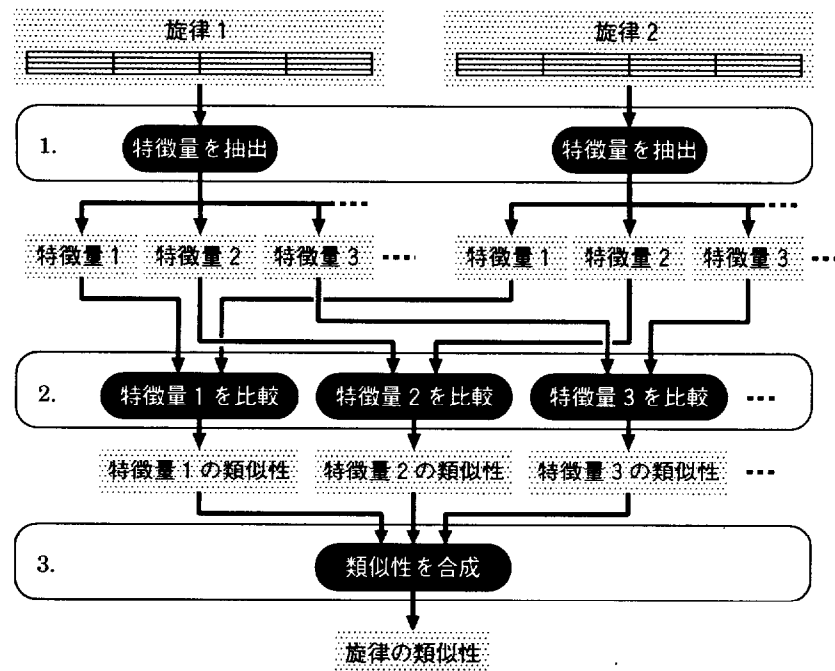


図 4.19: 旋律の類似性の評価手順

この手法では、それぞれの特徴量の数値が旋律断片の長さによらず同じような分布になるのであれば、その特徴量同士が類似している確率も旋律の長さに依らずほぼ一定の値になることが期待できる。これにより、旋律断片の長さに関わらず、旋律同士の類似確立がほぼ一定の高い値になる。従って、4.1.3 項で述べたように、長さが長い旋律断片に対する参考事例の検索に成功する確率も向上させることができる。

4.3.3 旋律の特徴量

本節で提案する旋律比較手法を用いるためには、どういった特徴量を利用するかが重要である。先にも述べたように、この手法で利用する特徴量は、以下の要件を満たしていなければならない。

- 旋律断片の長さによらずパラメータの数が一定
- それぞれの数値の分布が旋律断片の長さの影響を受けにくい

このような条件を満たす特徴量としては、以下のようなものが挙げられる。

- 平均音高

これは旋律断片に含まれる音の絶対音高の平均である。

- 平均音階

平均音階は音階名（ドレミファソラシド）とオクターブで表記した各音の音階の平均である。平均音高が絶対的な音の高さを表すのに対して、平均音階は調性の相違を排除したものである。

- 音高の偏差

音高の偏差は旋律断片に含まれる音のばらつきを表す。旋律断片に使われている音の幅を知ることができる。

- 音高差の平均

音高差の平均は旋律断片中の全ての音符について隣接する音との音高差を調べ平均したものである。この値が大きい旋律断片は音の動きが全般的に激しいことを表す。音高の偏差と組み合わせることで、旋律の音高の進行のより細かい特徴を捉えることが可能である。例えば、音高の偏差が大きく、音高差の平均が小さい場合、旋律断片の音高がゆるやかに変化していることがわかる。また、音高の偏差、音高差の平均が共に小さい旋律は全般に音の動きが小さい旋律である。

- 音階の各音の分布

音階の各音の分布は、12音階の各音階の出現頻度である。この分布は、長音階、短音階など、旋律断片内で使われているスケールの特性を表す特徴量である。また、旋律断片の長さやパートによってはコード的な特徴量を表すこともある。スケールは楽曲の雰囲気には大きな影響を与える情報であるため、この特徴量は旋律断片の雰囲気の比較に有効である。

- 平均音高の推移

平均音高の推移は、その旋律断片をいくつかに分割し、それぞれの間での平均音高差を表したものである。例として、旋律断片を二分割した場合について考えると、前半の平均音高に比べ後半の平均音高の方が高い場合、その旋律断片は上昇系の旋律である可能性が高い。逆に後半の平均音高の方が低い場合は下降系であると考えられる。

- 旋律断片の長さ

旋律断片の長さは、比較する旋律断片の総拍数である。如何に音階的特徴が類似していたとしても、その旋律断片の長さが極端に違う旋律ではあまり意味がないため、旋律断片の長さを考慮することも必要である。

- 音符の密度

音符の密度は旋律断片の中に含まれる音符の総数を、旋律断片の長さで割ったものである。

- 発音の密度

発音の密度は、音符の総数ではなく発音のある拍の総数を、旋律断片の長さで割ったものである。音符の密度では、和音があった場合に和音を構成している音符が全て勘定されるが、発音の密度では和音の構成音の数によらず、一つと勘定する。音符の密度が高い旋律断片は華やかなものが多く、発音の密度が高い楽曲はせわしない楽曲である。

- 拍ごとの音の分布

拍ごとの音の分布は、旋律断片内の各音符が小節内で何拍目に発音しているかを調べ、分布を取ったものである。一拍に集中している旋律断片はコードパートである可能性が高い。また、一小節の最終拍の頻度が高い旋律断片は、アウトタクトである可能性がある。

4.3.4 全体的な特徴の類似性の再帰的評価

前項で述べた演奏表情の生成手法は、単独で利用した場合、旋律の全体的な特徴しか判定できない。しかし、この手法を再帰的に利用することで子細な部分も考慮した類似性の評価を行なうことも可能である。ここではこの再帰的評価について簡単に説明する。

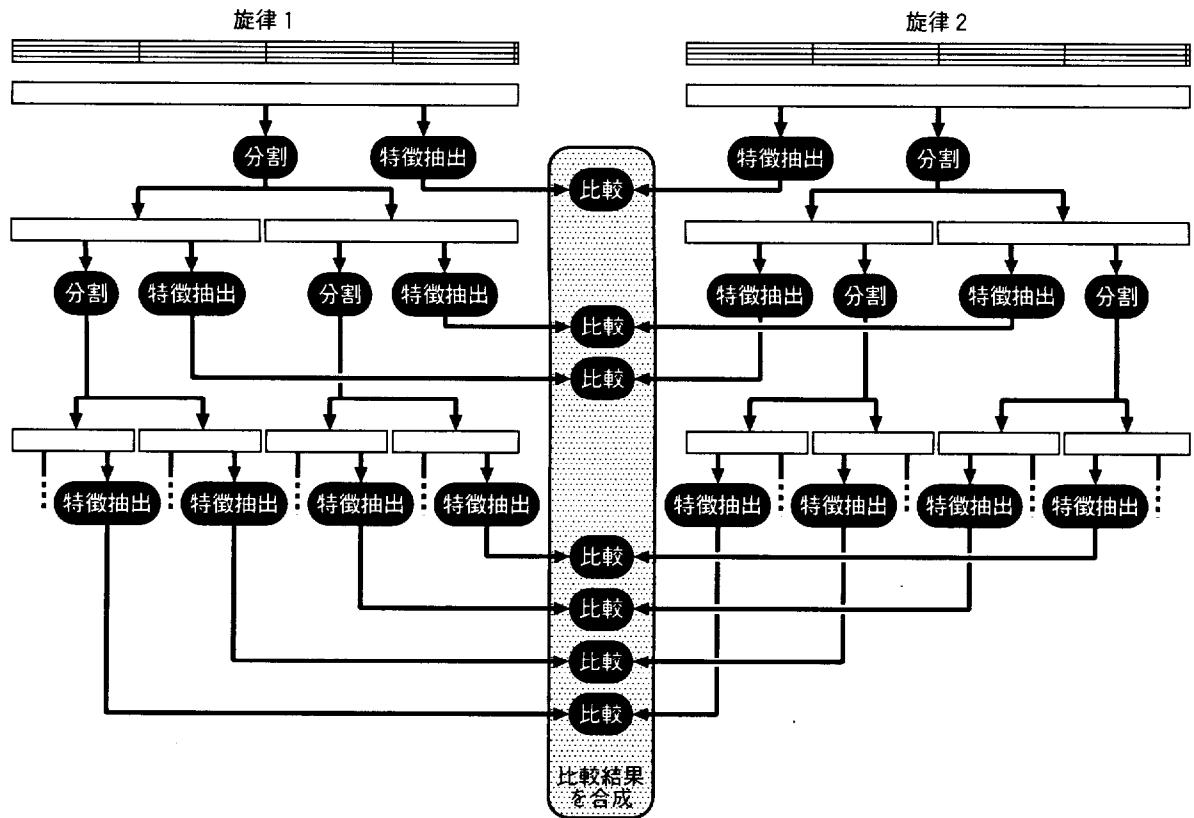


図 4.20: 旋律の全体的な特徴を利用した再帰的評価

この比較方法の手順を以下に示す（図 4.20）。

1. 比較する旋律断片を分割

比較する二つの旋律断片をそれぞれ同じ個数の小旋律断片に分割する。小旋律断片の分割数は 2 ~ 4 つ程度が適当である。分割された小旋律断片は交差していても構わない。分割の仕方は必ずしも長さの比率を同じにする必要はないが、それぞれの小旋律断片の長さが均等である方がよい。分割法は音楽的に根拠がある必要はなく、機械的に等分割しても問題はない。
2. 小旋律断片同士を比較

それぞれの対象断片の小旋律断片同士の全体的な特徴の類似性を評価する。類似性の評価は、先頭の小旋律断片から漸進的に行なう。
3. 結果を総合

旋律断片の全体的な類似性に小旋律断片の類似性を加味する。この計算は用途に応じて加算や乗算などを用いる。いずれの場合も、計算を行なう前に小旋律断片の類似性の正規化や重み付けを行なう。重みは評価の用途によって調整する。例えば、全体的な特徴を重視したい場合は、小旋律断片の類似性の重みを小さくする。逆に、詳細な部分の特徴を重視したい場合には、小旋律断片の類似性の重みを

大きくする。

4. 上記を再帰的に実行

個々の小旋律断片同士に対して、この手続きを再帰的に適用する。つまり、小旋律断片をさらにいくつかの微小旋律断片に分割し、それぞれの類似性を旋律断片の全体的な類似性に加味する。この繰り返しは、どちらかの旋律断片がこれ以上細かい小旋律断片に分割できなくなった時点で終了する。用途によっては、小旋律断片が一定の長さより短くなることを停止の条件としてもよい。例えば、一小節の拍数が異なる旋律断片を比較する場合は、小旋律断片の長さが小節より短くなったところで停止させる。

この比較手法は、長さが異なる旋律断片同士の比較が可能である。また、再帰の停止条件や小旋律断片の類似性への重みを調整することによって、類似性の評価結果の分布や特性を調整することが可能である。一方で、この手法はインデキシングが行ないにくいいため、データベース検索の用途には適さない。

4.3.5 旋律の類似性の評価例

ここでは旋律の類似性の比較方法の一例として、演奏表情生成システム Kagurame Phase-I (第5章) で用いる演奏表情の評価手法を譜例(図 4.21)を用いて説明する。なお、譜例では音部記号と拍子記号が省略されているが、いずれの楽曲もト音記号、四分の二拍子である。

旋律 1
音高 60 64 69 67 65 64 65 67 62 65 71 69 67 65 69

旋律 2
音高 62 71 69 71 67 69 66 67 64 73 71 73 69 71 69

図 4.21: 旋律の類似性の評価例 (ト音記号, 四分の二拍子)

4.3.5.1 特徴量

演奏表情生成システムでは、旋律断片の大局的な特徴量として、以下の3つを利用している。

- 主旋律の特徴
- 音階(スケール)の特徴
- リズムの特徴

まず、それぞれの特徴量について概説する。

主旋律の特徴 主旋律の特徴は、旋律断片の前半および後半の音符の平均音高の変化量である。この数値は主に旋律が上昇系であるか下降系であるかを表す特徴量である。式(4.30)に旋律断片 s の主旋律の特徴量 $C_m(s)$ の定義を示す。 N_f が旋律の前半の音符の集合、 N_l が旋律の後半の音符の集合、 $p(n)$ は音符 n

の音高を表す。主旋律の特徴の類似性は、式 (4.31) に示すように比較する二つの旋律断片の主旋律の特徴量の差の絶対値である。

$$C_m(s) = \frac{\sum_{n \in N_l} p(n)}{|N_l|} - \frac{\sum_{n \in N_f} p(n)}{|N_f|} \quad (4.30)$$

$$D_m(s_1, s_2) = |C_m(s_1) - C_m(s_2)| \quad (4.31)$$

図 4.21 の譜例の場合、いずれも 2 小節からなるので一小節目と二小節目の平均音高の差とする。旋律 1 の最初の C 音の音高を 60 とすると、旋律 1 (s_1) の前半の音高は (60, 64, 69, 67, 65, 64, 65, 67)、後半の音高は (62, 65, 71, 69, 67, 65, 67, 69) となる。従って音高の特徴量は式 (4.32) に示すように 1.75 である。同様に旋律 2 (s_2) は、前半及び後半の音高がそれぞれ (62, 71, 69, 71, 67, 69, 66, 67)、(64, 73, 71, 73, 69, 71, 67, 69) なので、音高の特徴量は 1.875 となる (式 (4.33))。よって主旋律の特徴量の類似性は 0.125 である (式 (4.34))。

$$C_m(s_1) = \frac{535}{8} - \frac{521}{8} = 1.75 \quad (4.32)$$

$$C_m(s_2) = \frac{557}{8} - \frac{542}{8} = 1.875 \quad (4.33)$$

$$D_m(s_1, s_2) = |1.75 - 1.875| = 0.125 \quad (4.34)$$

音階の特徴 音階の特徴は旋律断片の中の音符を階名で表した時のそれぞれの音の分布である。これは、旋律断片のスケールやコードなどのような特徴量を表す。

こでの階名とは、同じオクターブ内にある「ド」(楽曲の調において長音階を仮定した時の主音) との相対的な音高の差である。1 オクターブは 12 音であるため、この特徴量は 12 の要素を持つベクトルである。式 (4.35) と式 (4.36) に音階の特徴量ベクトルの計算式を示す。 N は旋律断片 s に含まれる音符の集合である。 $p(n)$ は n の音高、 $p(\text{do})$ はいずれかの「ド」音の音高を表す。この特徴量はスケールなどのような音の分布の特徴を表す。音階の特徴の類似性は、特徴量ベクトルの音階ごとの数値の差の総計である (式 (4.37))。

$$C_h = (c_{h,0} \ c_{h,1} \ \cdots \ c_{h,11}) \quad (4.35)$$

$$c_{h,i}(s) = \frac{|\{n | n \in N, ((p(n) - p(\text{do})) \bmod 12 = i)\}|}{|N|} \quad (4.36)$$

$$D_h(s_1, s_2) = \sum_{0 \leq i \leq 11} |c_{h,i}(s_1) - c_{h,i}(s_2)| \quad (4.37)$$

図 4.21 の旋律 1 はハ長調なのでド音の音高は 60, 72 などである。これに基づいて各音の音高を階名で表記すると (I, III, VI, V, IV, III, IV, V, II, IV, VII, VI, V, IV, V, VI) となる。旋律 2 の場合、ニ長調なのでド音は 62, 74 などとなる。それぞれの旋律の音階の分布および特徴量は表 4.3 に示す通りである。表中の #I, #II, #IV, #V, #VI は半音を表す。いずれの旋律も音符の総数が 16 なので、特徴量は分布の数値を 16 で割ったものである。音階の類似性は音階ごとの差の総和なので 0.25 である。

	分布		特徴量		
	s_1	s_2	s_1	s_2	$ s_1 - s_2 $
I	1	1	0.063	0.063	0.000
#I	0	0	0.000	0.000	0.000
II	1	1	0.063	0.063	0.000
#II	0	0	0.000	0.000	0.000
III	2	1	0.125	0.063	0.063
IV	4	3	0.250	0.188	0.063
#IV	0	0	0.000	0.000	0.000
V	4	4	0.250	0.250	0.000
#V	0	0	0.000	0.000	0.000
VI	3	4	0.188	0.250	0.063
#VI	0	0	0.000	0.000	0.000
VII	1	2	0.063	0.125	0.063
合計	16	16			0.25

表 4.3: 評価例の旋律の音階の分布と特徴量ベクトル

リズムの特徴 リズムの特徴は旋律断片の長さを拍単位で表したものである。あまりにも長さの異なる旋律断片は、主旋律や音階の特徴が類似していたとしても有効な事例とは言いがたいことから、旋律断片の類似性に旋律断片の長さを考慮する。リズムの特徴の類似性は、長い方の断片の拍数を短い方の断片の拍数で割ったものである（式 (4.38)）。 l_s は旋律断片 s の拍数である。

$$D_r(s_1, s_2) = \begin{cases} l(s_1)/l(s_2) & l(s_1) \geq l(s_2) \\ l(s_2)/l(s_1) & l(s_2) > l(s_1) \end{cases} \quad (4.38)$$

図 4.21 の旋律はいずれも四分の二拍子で長さが二小節なので拍数は 4 である。よって、類似性は 1 となる。

4.3.5.2 特徴量の合成

* 実験システムではこれらの特徴量の類似性を式 (4.39) で正規化し合計したものを旋律断片の類似性としている。期待値を揃えるため D_h を 6 倍し D_r から 1 引いている。図 4.21 の 2 旋律の場合、式 (4.40) に示すように類似性は 1.625 となる。

$$D = D_m + 6 \cdot D_h + (D_r - 1) \quad (4.39)$$

$$D(s_1, s_2) = 0.125 + 6 \cdot 0.25 + (1 - 1) = 1.625 \quad (4.40)$$

この数値は類似性が高いほど値が小さくなる。

* 演奏表情生成システムの類似旋律の検索では、この類似性の数値が低い旋律を参考事例として利用している。また、事例の重要度は類似性が高いものの数値が高くなる必要があるので、符合を反転し指数関数に適用したものの $e^{-D(s_1, s_2)}$ を重要度として利用している。

4.4 演奏状況

演奏表情は、演奏者や演奏のスタイル、演奏者の演奏時の演奏意図や気分、演奏時の疲労度、前後に演奏した楽曲など、様々な楽曲外の要因の影響を受けて大きく変化する。これらの演奏表情に影響を与える楽曲外の要因が演奏状況である。演奏状況に応じた演奏表情の生成を行なうためには、これらの演奏状況の計算機上での表現法について検討する必要がある。また、事例に基づく演奏表情生成手法では、参考事例の重要度の評価の際に演奏状況の類似性を考慮する必要がある。そこで、本節では演奏状況について概観した上で、演奏状況の記述方法、および演奏状況の類似性の評価方法について述べる。

4.4.1 演奏状況の表現

演奏状況は、先にも述べたように、演奏表情に影響を及ぼす楽曲外の要因である。まず、そのなかでも重要な要因とそれに関して我々が普段から用いている表現について概観する。代表的な演奏状況を以下に挙げる。

- 演奏者に関するもの

演奏状況のなかでも演奏者に関する情報は非常に重要である。演奏者に関する演奏状況は、基本的に演奏者が誰であるか、つまり、「演奏者が A 氏である」という情報で表現することができる。A 氏の演奏を良く知っている人であれば、これだけで演奏表情の特徴を特定することも可能であろう。しかし、演奏者の知名度が高くない場合など、演奏者の名前だけでは演奏者固有の演奏表情の特徴が十分に判断できないこともある。そこで、演奏者を補完する情報として、演奏者の経歴や趣向を付記することが多い。例えば、「B 氏と C 氏に師事した」や「D 音楽院出身」、「作曲家 E の演奏に精通する」といった情報である。

- 演奏のスタイル

演奏のスタイルは、演奏表情の与え方のパターンのようなものである。「ロマン派風」や「バロック風」、「古典風」のように、音楽史に基づくジャンル分けで表現する。これは演奏曲の歴史的背景に基づいて判断されることが多いことから、厳密には完全に楽曲外の情報であると断言することはできないが、演奏時に歴史的背景を無視した演奏スタイルで演奏を行なうことも可能であるため、ここでは演奏状況の一つとして考える。

- 演奏者の意図や気分

演奏者の意図や気分は、主に「明るく」や「はずむように」といった曲想表現を用いて表すことが多い。これは、演奏者や演奏曲が同じであっても演奏時の前後の文脈によって大きく変化するため、上記の二つほど明確ではない場合が多い。また、意図や気分に基づく要因の大部分は、その程度によって演奏表情に与える影響が変化する。例えば、単に「明るい」演奏といっても、その中にはただ明るい演奏だけではなく「非常に明るい」演奏や「やや明るい」演奏などの程度の違いがある。従って、この程度の大小も演奏状況を構成する要因の一つである。

これらを整理すると、我々は演奏者や演奏スタイルなどの要因ごとに、演奏者名やスタイル名などの具体的な属性を与えることで演奏状況を表現していることがわかる。また、演奏状況の要因によっては曲想表現のように、属性だけに程度を表す副詞や修飾句を合わせて表現している。

4.4.2 計算機での演奏状況の記述

ここでは、演奏状況の計算機上での表現法として、素性の集合による表現を提案する。素性の集合による表現では、演奏状況 C を素性 f の集合で表す (式 (4.41))。個々の素性は素性名 k と程度量 v の組合せで表現する (式 (4.42))。

$$C = \{ f \} \quad (4.41)$$

$$f = \langle k, v \rangle \quad (4.42)$$

演奏状況の素性とは上記のような演奏表情を決定する個々の要因である。素性名はその要因を表す名前前で、演奏状況として考えられる全ての要因 F のいずれかである (式 (4.43))。例えば「A 氏による明るくはずむような演奏」という演奏状況は、3つの素性からなり、それぞれ「A 氏が演奏」、「明るい」、「はずむような」という素性名を持つ。程度量は個々の素性の要因の程度を表す数値で -1 から 1 の間の実数値である (式 (4.44))。

$$k \in F \quad (4.43)$$

$$-1 \leq v \leq 1 \quad (4.44)$$

先に述べたように、演奏者の意図などの要因は「やや」とか「非常に」といった形容を伴っていることがある。この形容を数値で表現したものが程度量である。先の「A 氏による明るくはずむような演奏」の場合、いずれの要因も形容を伴っていないので、標準的な値であると考えられる。「明るい」や「はずむような」のような曲想表現を用いて表される素性については、形容を伴っていない場合は正の値の中間的な数として 0.5 を付与する。しかし、これは演奏者の意図の程度に応じて増減することもありうる。この演奏データが、ただ単に「明るい」ではなく「非常に明るい」演奏の場合は、「明るい」の程度量を 1 に近い値にし、逆に「若干明るい」程度であれば程度量を 0 に近づける。そして、「明るくない」といった否定的な素性には負の程度量を与える。一方で、「暗い」という素性があった場合、「明るい」という素性に負の程度量を与えることはせずに、「暗い」という素性に正の程度量を与える。これは、曲想によっては反意語が常に完全に反対の曲想を表すとは限らないためである。なお、「A 氏が演奏」などのように演奏時の状況や演奏者に関する事実を表す素性は、その事が明らかであれば形容を伴わなくても 1 に近い程度量を与える。演奏状況の例として「A 氏による明るくはずむような演奏」の演奏状況を表 4.4 に示す。

素性名	程度量
A 氏が演奏	1.0
明るい	0.5
はずむような	0.5

表 4.4: 「A 氏による明るくはずむような演奏」の演奏状況

素性には任意の要因をとり得るので、演奏データの演奏状況を厳密に定義したければ考えられる限りの素性を付与すればよい。しかし、次節で述べる演奏状況の比較では、比較する 2つの演奏状況の素性に一貫性がないと演奏状況の類似性の評価を行なうことができない。演奏表情生成時に与えた演奏状況の素性が演奏データ集に一度も登場していないと、当然ながらその演奏状況を反映した演奏表情の生成ができない。与えられた演奏状況の演奏データが本当に一つもないのであれば仕方がないが、同じような演奏状況の楽曲があるにも関わらず、素性の表記がたまたま違っていたために同じ素性のものがないというような事態は避けるべきである。

4.4.3 演奏状況の類似性の評価

我々の提案する演奏表情生成手法では参考事例の重要度を計算する際に入力された演奏状況と演奏データの演奏状況の類似性を評価する必要がある。ここでは、前項で述べた素性による演奏状況の記述を用いて、演奏状況の類似性を評価する手法について記す。

まず演奏状況空間および演奏状況ベクトルという概念を定義する。演奏状況空間は、演奏状況として考えられる全ての要因 F によって張られる多次元空間である。この空間の次数は要因の異なり数 $|F|$ に一致する。現実的には考えられる全ての要因をあらかじめ定義することは不可能なので、事例に基づく演奏表情生成手法の場合は、演奏データ集中で使われている演奏状況の素性の集合を F とする。例えば、「やや明るい演奏」、「明るくはずむような演奏」、「非常に重たく冷たい演奏」の3つの演奏データからなる演奏データ集の場合、演奏状況空間は「明るい」、「はずむような」、「重たい」、「冷たい」の4つの素性によって張られる4次元空間である。

そして、演奏状況の素性をそれぞれ演奏状況空間上の一次独立な単位ベクトルと考える。また、演奏状況の素性の程度量を単位ベクトルの係数と考える。これにより、任意の演奏状況は演奏状況ベクトルの一次結合で表すことができる。このベクトル \mathbf{v}_C を演奏状況ベクトルと呼ぶ。式 (4.45) に演奏状況 C の演奏状況ベクトル \mathbf{v}_C を示す。 k_f は素性 f の素性名、 v_f は f の程度量、 \mathbf{e}_{k_f} は素性名 k_f によって表される演奏状況の要因の単位ベクトルである。

$$\mathbf{v}_C = \sum_{f \in C} v_f \cdot \mathbf{e}_{k_f} \quad (4.45)$$

「A 氏による明るくはずむような演奏」という演奏状況 (表 4.4) の演奏状況ベクトルは、式 (4.46) のようになる。

$$\mathbf{v} = (1.0 \quad 0.5 \quad 0.5) \begin{pmatrix} \mathbf{e}_{A \text{氏が演奏}} \\ \mathbf{e}_{\text{明るい}} \\ \mathbf{e}_{\text{はずむような}} \end{pmatrix} \quad (4.46)$$

演奏状況で使われている素性 f はいずれも演奏状況空間を張る素性の集合 F に含まれることから、すべての演奏状況ベクトル \mathbf{v} は演奏状況空間上のベクトルとなる。それぞれの素性の程度量はその素性が顕著であればあるほど係数の絶対値が大きくなるので、特徴的な演奏は演奏状況空間上でのベクトルのノルムが大きくなり、また標準的な演奏は逆にノルムが小さくなる。

演奏状況が類似している場合、演奏状況に含まれる素性の種類やそれぞれの素性が持つ程度量の比率も近くなるので、演奏状況空間での演奏状況ベクトルの向きや大きさが近くなる。従って、比較する演奏状況ベクトルどうしの演奏状況空間上での向きや係数の数値の近さから演奏状況の類似性を評価することができる。式 (4.47)、式 (4.47) に演奏状況 C の演奏状況ベクトル \mathbf{v}_C と演奏状況 D の演奏状況ベクトル \mathbf{v}_D の類似性 $R(\mathbf{v}_C, \mathbf{v}_D)$ の定義を示す。

$$R(\mathbf{v}_C, \mathbf{v}_D) \equiv \frac{\mathbf{v}_C \cdot \mathbf{v}_D}{(\max\{|\mathbf{v}_C|, |\mathbf{v}_D|\})^2} \quad (4.47)$$

$$\equiv \frac{\sum_{f \in C \cap D} v_{f_C} \cdot v_{f_D}}{\max\{\sum_{f \in C} v_{f_C}^2, \sum_{f \in D} v_{f_D}^2\}} \quad (4.48)$$

式 (4.47) の分子 $\mathbf{v}_C \cdot \mathbf{v}_D$ は演奏状況ベクトルの内積を表す。また、 $|\mathbf{v}_C|$ は \mathbf{v}_C のノルムである。 $C \cap D$ は C と D の両方に含まれる素性の集合である。正確には $C \cap D$ ではなく $C \cup D$ が正しいが、 f が C もしくは D のどちらか片方にしか含まれていない場合、総和記号の右の値が 0 になるため、 $C \cap D$ でも同値である。 v_{f_C} は演奏状況 C に含まれる素性 f の程度量を表す。

内積をそのまま使用すると演奏状況ベクトルの大きさによる違いが考慮できないので、大きい方の演奏状況ベクトルのノルムの2乗で割って正規化する。例えば、内積をそのまま利用すると、非常に明るい演奏 $(1 \cdot e_{\text{明るい}})$ とやや明るい演奏 $(0.25 \cdot e_{\text{明るい}})$ の間の類似性と、明るい演奏 $(0.5 \cdot e_{\text{明るい}})$ 同士の類似性が共に 0.25 になるが、式 (4.47) の様に大きい方のノルムの二乗で割ると、それぞれが 0.25 と 1 になり、ベクトルの大きさが同じである後の方が高い値になる。

この定義では F 中のすべての素性 f を互いに独立な単位ベクトルとしているが、厳密には全ての素性 f が一次独立なわけではなく、中には互いに依存関係にある素性の組も少なくない。例えば、明るい演奏ははずむような演奏の特徴を含んでいることが多いし、逆に明るい演奏が同時に重たい演奏であるということはまれである。しかしながら、こういった素性間の依存関係や関連性を機械的に評価するのは困難なので、個々の素性は一次独立であると仮定している。

第5章 演奏表情生成システム

第3章で提案した事例に基づく演奏表情生成手法を利用し、演奏データ生成システム Kagurame Phase-I を実装した [61, 62]. この演奏データ生成システムは演奏曲の楽譜情報と演奏状況設定を入力として与えると、演奏曲に対して演奏状況に応じた演奏表情を生成し、その演奏データを出力するシステムである. Kagurame Phase-I は事例に基づく演奏表情生成手法の能力を検証することを主要な目的としている. Kagurame Phase-I を利用して演奏データを生成し、そのデータと人間の演奏との比較によって、事例に基づく演奏表情生成手法の有効性を評価した. また、同じ楽曲に対して様々な演奏状況設定を入力し、演奏状況への対応能力を検証した. 本章では、この演奏表情システム Kagurame Phase-I の概要、構成及び演奏表情の生成実験について述べる.

5.1 概要

5.1.1 Kagurame Phase-I の制約

このシステムでは事例に基づく演奏表情生成手法の有効性を検証することを主要な目的としており、システムの構成を単純にするために次にあげる制約がある.

- 対象曲は単旋律の楽曲のみに限定する
- 楽曲の境界は人間が決定し入力として与える
- 弾き損じや弾き間違いにより楽譜と異なる演奏が行なわれた演奏データは扱わない

単旋律とは、和音を含まないという意味である. 和音を含む楽曲は、単旋律の楽曲に比べて様々な複雑な要素がある. まず、参考事例の検索や重要度の計算の際に和音の組合せの類似性も考慮しなければならないため、旋律の類似性の評価が複雑になる. また、取り扱うべき演奏表情の種類も増える. 例えば、和音を含む楽曲では、同時に鳴るべき音のずれや和音を構成する音の間の音の強さのバランスを考慮する必要がある. これらの複雑化の要因は、事例に基づく演奏表情生成手法の能力を評価する際に妨げになる可能性があるため、Kagurame Phase-I では単旋律の楽曲のみを対象とする.

旋律断片の分割を行なうためには、楽曲構成情報が必要となる. 楽曲構成情報は 4.1.4 項で述べたように、主題、繰り返し、フレーズ、小節などの楽曲を構成する部品の情報である. これらの情報の大部分は、楽譜から直接的、あるいは間接的に取得することが可能であるが、そのためには楽譜に記載されている音符以外の記号の情報を入力する必要がある. また、旋律断片の分割が妥当ではなかった場合に、演奏表情の生成の妨げとなる可能性がある. そこで、Kagurame Phase-I では旋律断片の境界の決定をあらかじめ人手で行ない、楽譜情報と共に付与する.

弾き損じとは、本来楽譜に書かれている音符とは違う音符を打鍵してしまうことをさす. 本手法で利用する演奏データは、楽譜上に書かれている音符を過不足なく順番通りに演奏したものである必要がある. 音符が欠落している場合や余計な音符が混入している場合、その前後の演奏表情が演奏者の意図していた演奏表情とは異なったものであることが多い. 例えば、演奏者が弾いた音符を間違え、その音符を弾き直した場合、その区間の局所テンポが極端に遅くなる傾向がある. また、音符の欠落などは演奏データと楽

譜との対応づけや局所テンポの計算などの処理の妨げとなる。従って、Kagurame Phase-I では弾き損じや弾き間違いが含まれる演奏データは使用できない。

5.1.2 システムが取り扱う情報の種別と表記法

Kagurame Phase-I で取り扱う情報は以下の四種類である。

- 演奏情報
- 楽譜情報
- 楽曲構成境界情報
- 演奏状況情報

演奏情報は、演奏表情を伴った演奏の記録である。これは、演奏データ集やシステムの出力で使っている。楽譜情報は、楽曲の音符情報などの楽譜に記載されている情報で、演奏データ集と入力で使っている。楽曲構成境界情報は、楽曲の旋律断片の境界の決定に使用する情報である。これは、演奏データ集と入力に於いて楽譜情報と一緒に付与するものである。演奏状況情報は、演奏データごとの演奏状況の記録や、システムに入力する演奏状況設定で使用する。これらの内、楽譜情報と楽曲構成境界情報は同じファイルに併記している。

ここでは、それぞれの情報の内容や書式について概説する。

演奏情報 演奏情報は、個々の音符の演奏時刻を時間順に記録した情報である。Kagurame Phase-I では演奏情報の記録に標準 MIDI フォーマット (SMF) [54] のシーケンスファイルを使用する。SMF は打鍵や離鍵など演奏時の演奏動作を時間順に記録したものである。

MIDI とは、シンセサイザや計算機などの間で演奏に関する情報を交換するためのプロトコルである [5]。MIDI ではケーブルや端子などの物理層の構成から、転送するデータの書式や意味が定められている。近年は殆んどすべての電子楽器に MIDI 端子と呼ばれるコネクタが装備されている。この MIDI 端子同士をケーブルで接続し MIDI 信号を送信することによって、他の電子楽器を発音させるなどのコントロールを行なうことが可能である。

MIDI 端子を持った電子楽器に対して演奏操作が行なわれると、ほぼ同時にその操作内容が MIDI 信号として送出する。鍵盤楽器の場合は打鍵や離鍵、ペダル操作などがある。これらの演奏操作を MIDI イベントと呼ぶ。MIDI では発音操作の開始と終了を別個に取り扱うため、鍵盤楽器の場合は打鍵情報と離鍵情報が別々に送られてくる。MIDI 信号の発音開始情報（鍵盤楽器の場合は打鍵情報）は発音された音の音高と発音の強さからなる。いずれも 0 から 127 の間の数値である。音高は A3（440Hz の音）が 57 で、半音ごとに 1 上下する。平均率を前提としているため、G#4 と Ab4 などの様に鍵盤上同じキーになるものは全て同じ値である。なお、MIDI では発音の強さをベロシティと呼んでいる。

計算機用の MIDI 接続装置も安価で市販されているため、計算機と電子楽器を MIDI を介して接続することができる。従って、電子楽器の MIDI 出力を計算機の MIDI 入力に接続し、これらの MIDI 信号を受信することで、演奏操作とほぼ同時にその操作の内容を計算機側で知ることが可能である。また、計算機で受信した MIDI イベントを記録することによって、電子楽器で行なった演奏を保存することも可能である。MIDI 信号には時間情報が含まれていないので、MIDI 信号の受信時に計算機側で内部時計を参照し、その時刻を演奏情報と一緒に記録する必要がある。このようにして記録された演奏情報がシーケンスファイルである。SMF はシーケンスデータの標準的なフォーマットで、市販のアプリケーションでも作成することが可能である。

SMF は、発生した MIDI イベントを時間順に羅列したものである。SMF 中のそれぞれの MIDI イベントは以下の 3 つの情報からなる。

- デルタタイム
- イベントコード
- 付加情報

デルタタイムは直前の MIDI イベントとの時間間隔である。デルタタイムの単位は SMF ごとに様々で、一般的には一拍を 48 や 96 と定めた拍単位の値や、1 を 10 ミリ秒程度に設定した時間単位の値で与えられることが多い。イベントコードは打鍵や離鍵などそのイベントの種別を表す。付加情報はイベントを補足する情報で、打鍵イベントの場合は音高と音の強さ、離鍵イベントの場合は音高がそれにあたる。



図 5.1: 図 5.2 の SMF の楽譜

```
00 92 30 60 00 92 3C 60
60 91 43 40 60 90 4C 20
81 40 82 30 00 00 82 3C
00 00 81 43 00 00 80 4C
00 ...
```

図 5.2: SMF の例

デルタタイム	イベントコード	付加情報	注釈
00 (0)	92 (打鍵)	30 (48), 60 (96)	時刻 0 に音高 48 (C3) を強さ 96 で打鍵
00 (0)	92 (打鍵)	3C (60), 60 (96)	時刻 0 に音高 60 (C4) を強さ 96 で打鍵
60 (96)	91 (打鍵)	43 (67), 40 (64)	時刻 96 に音高 67 (G4) を強さ 64 で打鍵
60 (96)	90 (打鍵)	4C (76), 20 (32)	時刻 192 に音高 76 (E5) を強さ 32 で打鍵
81 40 (192)	82 (離鍵)	30 (48), 00 (0)	時刻 384 に音高 48 (C3) を離鍵
0 (0)	82 (離鍵)	3C (60), 00 (0)	時刻 384 に音高 60 (C4) を離鍵
0 (0)	81 (離鍵)	43 (67), 00 (0)	時刻 384 に音高 67 (G4) を離鍵
0 (0)	80 (離鍵)	4C (76), 00 (0)	時刻 384 に音高 76 (E5) を離鍵
	⋮		⋮

表 5.1: SMF の例

図 5.2, 表 5.1 に SMF の例を示す。図 5.2 は図 5.1 に示す楽譜を記した SMF の内容である。表 5.1 は、この SMF 中にある個々の MIDI イベントについて解説したものである。表 5.1 に示すように、個々の音符の打鍵及び離鍵などの MIDI イベントがその発生順に記録されている。このデータのデルタタイムの値は、四分音符の長さを 96 とした拍単位の長さである。このデータは楽譜情報からツールを使って作成したものであるため、デルタタイムが拍単位で表されており、音符の拍位置や演奏長も楽譜に示された比率を正確に再現している。一方で、人間の演奏データでは、デルタタイムを時間単位で表すことが多く、また、各音符の拍位置や演奏長もテンポなどの揺らぎに応じて、楽譜からずれた不均等な値となる。

Kagurame Phase-I では、演奏データ集の人間の演奏と、出力する演奏データの記録フォーマットとして SMF を利用している。

楽譜情報 楽譜情報は楽譜上に記載された楽曲に関する情報である。楽譜上には様々な記号が記載されているが、Kagurame Phase-I ではそのなかから、以下の二つを楽譜情報として与える。

- 楽曲の調
- 音符情報

強弱記号やテンポ表示、繰り返し記号など、楽譜に記載されているその他の情報は付与しない。

楽曲の調は楽曲の主音の音高（ド音）を 1 ないしは 2 文字の文字列で記したものである。表記を表 5.2 に示す。補助記号の b はフラット（半音階下げる）を意味する。

主記号	A, B, C, D, E, F, G
補助記号	なし, #, b

表 5.2: 調の表記

主音の音高は楽譜の左端に記された調号から決定する。長調と短調の区別は行なわず、全て長調であるものとする。従って、短調の楽曲でも調の表記はド音の音高から決定する。なお、Kagurame Phase-I では純正率ではなく平均率を想定しているため、A# と Bb などの様に、主音の音高が同じになる調は全て同じ調を表す。

音符情報は以下の情報からなる。楽曲中に登場する全ての音符についてこの 4 種の情報を記録する。

- 拍位置
- 音譜長 (拍単位)
- 音高 (階名)
- オクターブ

拍位置は、音符の楽譜上での位置を表す。楽曲の先頭からの拍数で表記する。一拍の長さは楽譜にある拍子記号に従う。拍子記号の下の方が一拍の分数である。例えば、四分の四拍子や四分の三拍子の場合是一拍の長さが四分音符、八分の六拍子の場合八分音符となる。拍位置が拍頭以外の場合は小数で表記する。

音譜長は、音符の楽譜上での長さを拍数で表現したものである。一拍の長さは拍位置と同様に拍子記号から決定する。また、一拍の長さより短い場合は小数で表す。一拍の長さが四分音符の場合、全音符が 4、八分音符が 0.5、付点四分音符が 1.5 となる。

階名	表記
ド	0
ド#	0.5
レ	1
レ#	1.5
ミ	2
ファ	3
ファ#	3.5
ソ	4
ソ#	4.5
ラ	5
ラ#	5.5
シ	6

表 5.3: 長調の場合の階名の表記

音高は音名ではなく階名に基づいた表記を使用する。階名とは、楽曲の主音との相対的な音高差である。具体的には、ド音を 0 とし、レを 1、ミを 2 という様に表記する。半音は 0.5 で表す。表 5.3 に長調の場合の階名と数値との関係を示す。ミとシは半音がないため 2.5 と 6.5 は使われない。

オクターブはその音符が属する音階を表す。オクターブ値は C4 から B4 (440Hz の音を A3 とする) の間に主音があるオクターブを 5 とする。調ごとのオクターブ値の範囲を図 5.3 に示す。

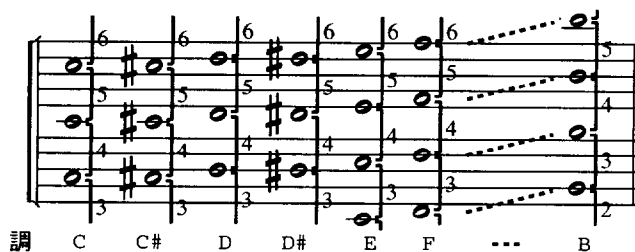


図 5.3: 調, 音高, オクターブ値の関係 () の右側の数値がオクターブ値)

SMF で用いられている絶対音高は、楽曲の調、音符の音階とオクターブから計算することができる。この計算を式 (5.1) に示す。

$$\text{絶対音高} = \text{基本音高} + \text{オクターブ} * 12 + \text{階名の音高} \quad (5.1)$$

式 (5.1) 中の「基本音高」と「階名の音高」は、それぞれ表 5.4 と表 5.5 に示す値である。例えば、へ長調 (F) の楽曲中の、オクターブ 5、音高 2 (ミ音) の音符の音高は、 $5 + 5 * 12 + 4 = 57$ である。

先にも述べたように、楽譜情報は楽曲構成情報と共に単一のファイルで与えている。この書式は楽曲構成情報の説明中で述べる。

楽曲構成境界情報 楽曲構成情報は 4.1.4 項 で述べた、楽曲を構成する部品の情報である。Kagurame Phase-I では、旋律断片ごとにその子となる旋律断片の組合せを記述し、楽曲の構成情報を記す。書式を

調の表記	音高の基本値
C, Bb	0
C#, Db	1
D	2
D#, Eb	3
E, Fb	4
F, E#	5
F#, Gb	6
G	7
G#, Ab	8
A	9
A#, Bb	10
B, Cb	11

表 5.4: 調ごとの基本音高

階名表記	音高の差分
0	0
0.5	1
1	2
1.5	3
2	4
3	5
3.5	6
4	7
4.5	8
5	9
5.5	10
6	11

表 5.5: 階名と音高の対応

式 (5.2), 式 (5.3), 式 (5.4) に示す。旋律断片は式 (5.2) の書式 [59] で定義する。「旋律断片名」は、ここで定義する旋律断片を一意に区別するための名前であり、同じ楽曲の中に重複するものがあることはならない。「基準拍数」と「基準拍長」はその旋律断片の長さを表す数値である。基準拍長はその旋律断片の中での基準となる長さを拍数で表したものである。基準拍数はその旋律断片の長さが基準拍長いくつ分であることを示す。よって、旋律断片の拍数は 基準拍数 × 基準拍長 である。旋律断片に属する子は「子の情報」で記す。子の情報はその旋律断片に含まれる子の数だけ列挙する。

子の情報の書式を式 (5.3), 式 (5.4) に示す。式 (5.3) は子が旋律断片の場合の書式である。「位置」はこの子が親となる旋律断片のどの位置から始まっているかを表す数値で、親となる旋律断片の基準拍長を 1 とする。この子が旋律断片の先頭に位置する場合は、「位置」を 1 と記す。「子の旋律断片名」はこの位置に配置されている旋律断片の名前である。この旋律断片は、他の部分で定義されていなければならない。

式 (5.4) は子が音符の場合の書式である。「位置」は子が旋律断片の場合と同様に、親となる旋律断片のどの位置にこの音符があるかを定義するものである。「拍長」は音符の長さを拍単位で記した数値である。オクターブ及び音高は先に述べた音符の表記法に準ずる。

旋律断片 := elm 旋律断片名(基準拍数/基準拍長) { 子の情報 + } (5.2)

子の情報 := 位置:子の旋律断片名() ""; (5.3)

|| 位置:Play(length:拍長, octave:オクターブ, note:音高) ""; (5.4)

図 5.4 に旋律断片の記述例を示す。この図の上の例で定義している旋律断片 (Segment1) は、2 つの旋律断片を子に持つ旋律断片である。この旋律断片は基準拍数と基準拍長が共に 2 なので、この旋律断片の長さは 4 拍である。1 拍目 (旋律断片冒頭) に Child1 という旋律断片、3 拍目に Child2 という旋律断片がある。

下の例 (Segment2) は、2 つの音符を子に持つ旋律断片である。この旋律断片は基準拍数が 2 で基準拍長 0.25 である。従って、旋律断片の長さは 0.5 拍である。旋律断片の冒頭に、長さが 0.25 拍のオクターブ 6 のミの音符がある。また、旋律断片の冒頭から 0.25 拍の位置に、長さが 0.25 のオクターブ 5 のソの音がある。

```
elm Segment1(2/2){
  1:Child1() "";
  2:Child2() "";
}
,
elm Segment2(2/0.25){
  1:Play(length:0.25, octave:6, note:3) "";
  2:Play(length:0.25, octave:5, note:5) "";
}
```

図 5.4: 楽譜情報と楽曲構成情報の例

演奏状況情報 演奏状況情報は素性名と程度量の組、もしくは素性名のみを列挙して表す。素性名のみの場合には程度量を 1 とみなす。図 5.5 の表記は「古典派風 (classical)」が 0.2 で「ロマン派風 (romantic)」が 1.0 であることを表す。よって、これは「基本的にロマン派風でやや古典派風がかかった演奏」という演奏状況を表している。

図 5.5: 演奏状況の記述例

5.1.3 Kagurame Phase-I で取り扱う演奏表情

Kagurame Phase-I では以下の三種類の演奏表情を取り扱う。

- 局所テンポ
- 音の強さ
- 演奏長比

それぞれの演奏表情値について以下で概説する。

テンポ 音符 i での局所テンポは一音ごとのテンポのことである。 n 個の音符からなる楽曲の i ($1 \leq i \leq n$) 番目の音符の局所テンポ T_i を式 (5.5) に示す。

$$T_i = \frac{b_{i+1} - b_i}{t_{i+1}^{on} - t_i^{on}} \quad (5.5)$$

b_i は i 番目の音符の拍位置である。 t_i^{on} は i 番目の音符が演奏された時刻を表す。鍵盤楽器の場合、 t_i^{on} は演奏者が鍵盤を叩いた時刻である。演奏データでは、SMF に記録されているノートオンイベントの時刻をそのまま利用する。拍位置および演奏時刻は曲の先頭からの絶対位置で表す。Kagurame Phase-I では時刻の単位を秒としているが、この単位を分とすれば、楽譜などで用いられている標準的なテンポ値となる。4.2.1 項で述べたように、テンポの値をそのまま計算すると平均などの計算によって誤差が生じるため、実際の計算には一拍の演奏時間 U_i を用いる。一拍の演奏時間 U_i は式 (5.6) に示すようにテンポの逆数である。

$$U_i = T_i^{-1} = \frac{t_{i+1}^{on} - t_i^{on}}{b_{i+1} - b_i} \quad (5.6)$$

Kagurame Phase-I では単旋律の曲を対象としているので、複数の音符が同じ時刻に演奏されることはない。従って b_{i+1} と b_i は必ず異なる値になる。よって、一拍の演奏時間の分母が 0 になることはない。

テンポや一拍の演奏時間を計算するためには i 番目と $i+1$ 番目の音符の演奏時刻が必要である。一方で、楽曲の最後の音符 n には $i+1$ に相当する音符がないため、これらの計算ができない。そこで、楽曲の最後の音符のテンポ U_n 及び一拍の演奏時間 T_n は、便宜的に、前の音符 $n-1$ の値と同じ値を利用する。

$$T_n = T_{n-1} = \frac{b_n - b_{n-1}}{t_n^{on} - t_{n-1}^{on}} \quad (5.7)$$

$$U_n = U_{n-1} = \frac{t_n^{on} - t_{n-1}^{on}}{b_n - b_{n-1}} \quad (5.8)$$

音の強さ 音の強さは MIDI のベロシティ値を利用している。ベロシティ値は MIDI に於いて打鍵の強さを表す数値である。この値は、0 から 127 までの数値である。MIDI では打鍵情報を音高とベロシティ値の組で表しているため、必ず一音ごとにベロシティ値を取得することができる。演奏データで利用されている SMF にも一音ごとにベロシティ値が記録されている。また、MIDI 楽器で演奏を行なう際にも演奏時に送信するベロシティ値によって音の強さをコントロールすることができる。

演奏長比 演奏長比は楽譜上に記載された音符の長さ与实际に演奏が行なわれた長さの比率を表す(図 5.6)。演奏長比が 1 の場合、楽譜通りの正確な演奏となる。通常は楽譜上の譜長に比べて若干短くなるため 1 より若干低い値になる。スタッカートのように、実際の譜長に比べて非常に短い演奏の場合 0 に近い数値になる。レガートなどの奏法では 1 を越えることもある(図 5.7)。

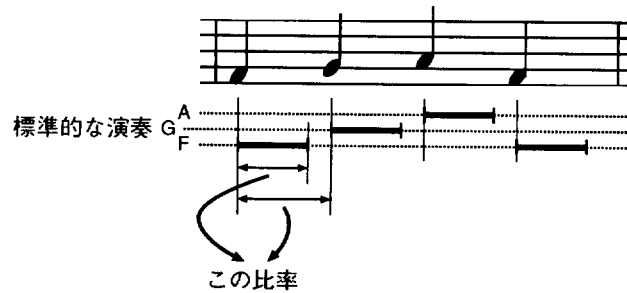


図 5.6: 演奏長比

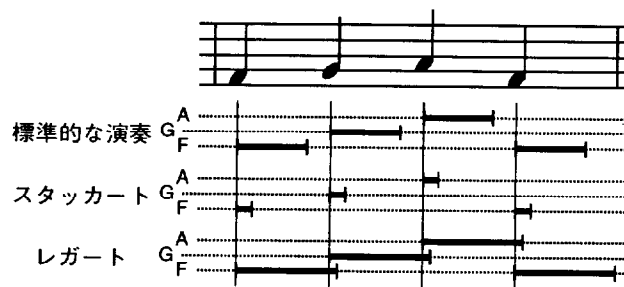


図 5.7: 演奏長比と演奏表情の関係

i 番目の音符の演奏長比 D_i を式 (5.9) に示す。

$$D_i = \frac{t_i^{off} - t_i^{on}}{l_i} \cdot T_i \quad (5.9)$$

l_i は音符 i の楽譜上での長さを拍数で表したものである。4 分の 4 拍子の場合、4 分音符が 1、8 分音符が 0.5 となる。 t_i^{on} は音符 i が演奏された時刻で、 t_i^{off} は音符 i の演奏が終了した時刻である。演奏データでは、 t_i^{on} および t_i^{off} は、それぞれ SMF に記録された i 番目の音符のノートオンイベント及びノートオフイベントの時刻をそのまま利用する。

鍵盤楽器の場合は、離鍵の時刻が t_i^{off} となる。また、弦楽器や管楽器では弦の摩擦や息の吹き込みなどの発音行為を停止した時刻である。

5.2 構成

本節では、Kagurame Phase-I の構成と演奏データの生成手順について概説する。

図 5.8 に Kagurame Phase-I 全体構成を示す。基本的な構成は 3.1.1 項に示した事例に基づく演奏表情生成手法と同じである。

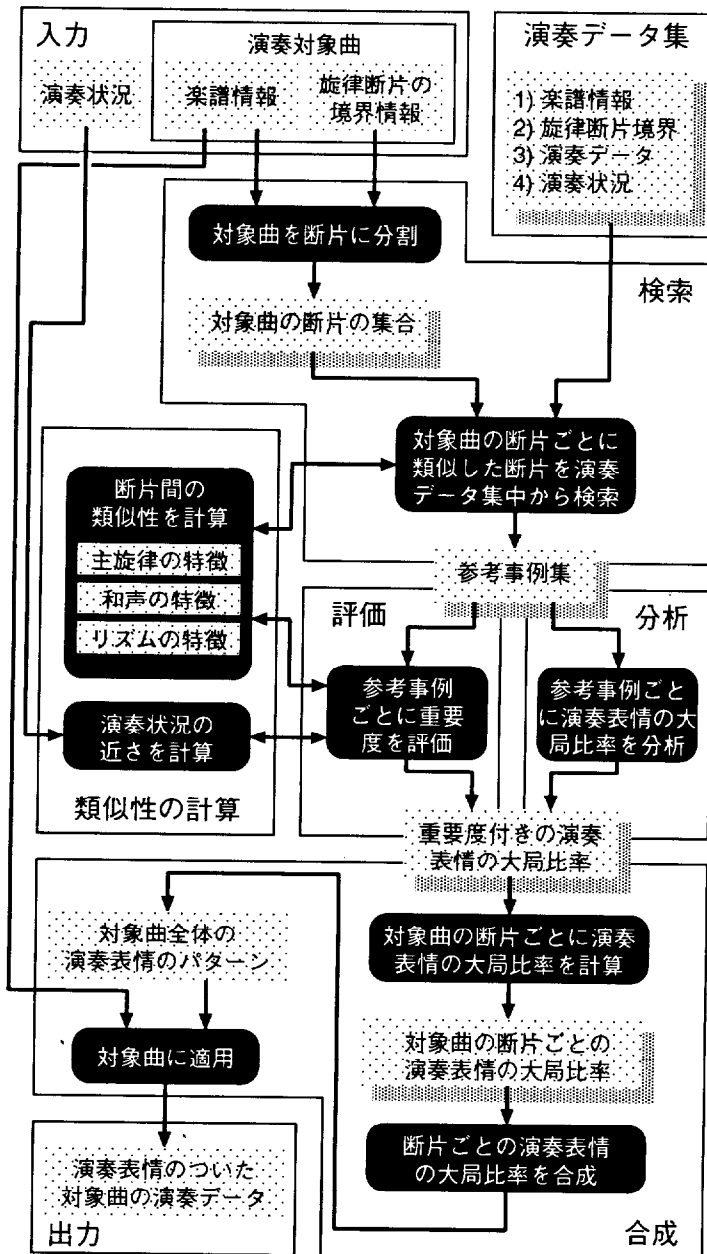


図 5.8: Kagurame Phase-I の全体構成

5.2.1 入出力と演奏データ

まず Kagurame Phase-I で利用するデータについて概説する。

演奏データ集 Kagurame Phase-I では演奏表情に関する知識として演奏データ集をあらかじめ用意する必要がある。演奏データ集の中の個々の演奏データはいずれも以下の 4 種類の情報からなる。

- 演奏曲の楽譜情報
- 演奏曲の旋律断片の境界
- 演奏データ
- 演奏データ収録時の演奏状況

これらの情報は 5.1.2 項で述べた形式で与える。このうち、演奏曲の楽譜情報と旋律断片の境界情報に関しては、同じ楽曲に対する演奏データが複数ある場合に、それらの間で同じファイルを共有することが可能である。

入力 Kagurame Phase-I への入力は以下の 3 種類の情報からなる。

- 演奏対象曲の楽譜情報
- 演奏対象曲の旋律断片の境界
- 生成する演奏データの演奏状況設定

いずれも 5.1.2 項で述べた形式で付与する。

出力 生成された演奏データは SMF ファイルで出力する。

5.2.2 演奏表情の生成手順

次に、Kagurame Phase-I の演奏表情生成手順を詳説する。Kagurame Phase-I の構成は 3.1.2 項に示した事例に基づく演奏表情生成手法の基本構成に準ずるが、付随する処理がいくつか追加されている。処理の手順を以下に示す。

1. 準備
 - (a) 演奏データ集の読み込み
 - (b) 入力の読み込み
2. 演奏表情生成
 - (a) 参考事例の検索
 - (b) 参考事例の重要度の評価
 - (c) 参考事例の演奏表情の分析
 - (d) 演奏対象曲の演奏表情の合成
 - (e) 演奏対象曲への演奏表情の適用
3. 出力
 - (a) 演奏表情情報を SMF に変換し出力

演奏データの読み込み まず最初に、前処理として演奏データ集の読み込みを行なう。演奏データの個々のデータは、先に述べたように、楽譜情報、境界情報、演奏データ、演奏状況の4つの情報からなる。このうち、楽曲情報と境界情報は同じファイルに記されているが、それ以外は別々に保存されている。演奏データの読み込みではそれぞれのファイルから読み込んだデータの一つにまとめる作業も行なう。演奏データの読み込み手順を以下に示す。

1. 楽譜情報と境界情報の読み込み
2. 演奏データの読み込み
3. 楽譜情報と演奏データとの対応づけ
4. 演奏状況の読み込み

楽譜情報と境界情報を記述したファイルを読み込む。それに続いて、演奏データを記録した SMF を読み込む。この後で、楽譜情報と演奏データを照らし合わせて、演奏データ中の個々のイベントが楽譜上のどの音符に対応するかを調査する。

SMF の解説で述べたように、SMF には同じ音符の打鍵と離鍵が別々のイベントとして記録されている。そこで、まず、SMF 中の打鍵イベントごとに対応する離鍵イベントを捜し出し、打鍵と離鍵のペアを作成する。続いて、楽譜と SMF データ中の打鍵イベントを先頭から順番に照らし合わせ、SMF 中のそれぞれのイベントが楽譜上のどの音符に対応するかを決定する（図 5.9）。

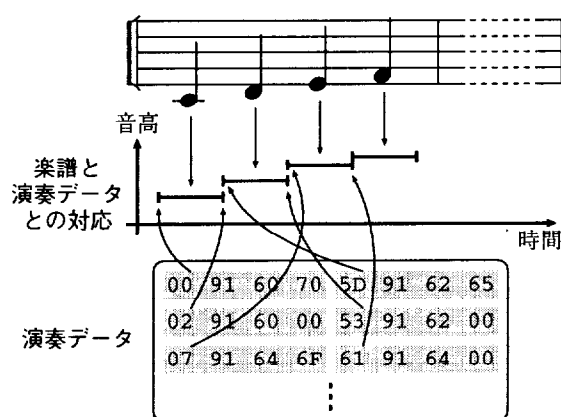


図 5.9: 演奏データと楽譜情報との対応づけ

最後に演奏データの演奏状況を読み込み、この情報をそれぞれの旋律断片に付加する。

この一連の手続きを全ての演奏データについて行なう。この結果、大局比率、楽譜情報、演奏状況を伴った旋律断片の集合が得られる。

入力を読み込み 入力、演奏曲の楽譜情報と境界情報、そして演奏状況の設定である。楽譜情報と境界情報は単一のファイルに記録されているのでそれを読み込む。これに演奏状況設定を付け加えて入力の処理は終了である。

参考事例の検索 参考事例の検索では、対象曲の旋律断片ごとに、演奏データ集に含まれる全ての旋律断片との間で類似性の評価を行ない、類似度が高かった旋律断片を参考事例として収集する。類似旋律の検索では以下の三つの特徴量を考慮している。

- 平均音高の差

- 音階の分布
- 拍数

平均音高の差は、旋律断片の前半と後半の音高にどの程度の差があるかを表している。これは旋律断片の全体的な傾向が上昇系か下降系かといった旋律の外形的な特徴を表す。音階の分布は旋律断片に含まれる音符の音階ごとの分布である。これにより、旋律断片が持つスケールの特徴がわかる。拍数は旋律断片の総拍数の比である。長さが極端に異なる旋律は、仮に上記の二つの特徴が似ていたとしても参考事例として適さないので、拍数の相違を考慮している。

Kagurame Phase-I で用いている旋律断片の類似性の評価アルゴリズムの詳細は、4.3.5 項 で述べた通りである。この結果を、式 (5.10) に示す指数関数に適用した数値が、一定の閾値を越えた旋律断片を参考事例として利用している。

$$e^{-\text{類似性}} > \text{閾値} \quad (5.10)$$

5.3 節で述べる実験では、閾値を 0.00001 に設定した。

参考事例の重要度の評価 事例の重要度は、以下の情報から決定する。

- 旋律断片同士の類似性
- 祖先の旋律断片の再帰的類似性
- 前後の旋律断片の類似性
- 演奏状況の類似性

旋律断片の類似性の評価アルゴリズムは 4.3.5 項 で述べた通りである。また、演奏状況の類似性の評価アルゴリズムは 4.4.3 項 で述べたように演奏状況ベクトルの内積から計算する。これらの類似性の数値を全て加算し、参考事例の検索と同様にその結果を式 (5.10) に示す指数関数に適用した。指数関数を適用する理由は主に二つある。一つは、類似性の数値の大小関係である。類似性の数値は、いずれも、最も似ているものが 0 で類似性が低くなるにつれて数値が大きくなる。一方で、重要度の数値はこれ以降の合成操作で加重平均の重みとして利用されるため、似ているものほど数値が大きくなる必要がある。そこで、指数関数の肩の類似性の数値の符号を負にし、類似性の数値が 0 に近いほど大きい値をとるようにしている。もう一つの理由は、類似性の数値の分布である。類似性の評価結果は、全く同じ旋律と全然似ていない旋律の数値の比がそれほど大きくなく、また数値の分布がほぼ均一である。従って、このままの数値を利用すると、それほど似ていない事例が加重平均で必要以上に影響を及ぼしてしまう。そこで、類似性が高いものほど分布が少なくなるようにするため、指数関数を適用している。

参考事例の演奏表情の分析 演奏表情の分析では、参考事例集の個々の事例について演奏表情の大局比率を計算する。手順を以下に示す。

1. 旋律断片の局所的な演奏表情値の計算
2. 演奏表情の大局比率の計算

まず、参考事例の旋律断片とその親について、テンポ、音の強さ、演奏長比の三種類の演奏表情の局所的な演奏表情値を算出する。この計算法は 5.1.3 項 に述べた通りである。そして、4.2.2 項で述べた手法に従って、参考事例の演奏表情の大局比率を算出し記録する。

この一連の計算を全ての参考事例に対して行なう。

演奏対象曲の演奏表情の合成 演奏表情の合成では、演奏対象曲の旋律断片ごとに、参考事例の大局比率を事例の重要度で加重平均し、旋律断片の大局比率を算出する。この計算は 4.2.3 項で述べた通りである。テンポ、音の強さ、演奏長比のそれぞれについてこの計算を行なう。

演奏対象曲への演奏表情の適用 合成処理によって得られた、各旋律断片の演奏状況の大局比率から楽曲の各音符の一音ごとの演奏表情値の復元を行なう。詳細は 4.2.3 項で述べた通りである。ここでも、テンポ、音の強さ、演奏長比のそれぞれについてこの計算を行なう。また、この計算で必要になる曲全体での演奏表情の平均値は、曲全体に対する参考事例の平均値を事例の重要度で加重平均して算出する。

演奏データの出力 演奏対象曲に演奏表情を適用した結果得られる演奏対象曲の一音ごとの演奏表情値から、演奏対象曲の演奏表情付きの演奏データを作成し SMF に出力する。この処理は、主に以下の 3 つの手続きで構成されている。

1. 打鍵時刻の計算
2. 離鍵時刻の計算
3. データの出力

まず、音符の局所テンポから音符の打鍵時刻を算出する。SMF は発音時刻を直前のイベントとの相対的な時間差で記述するため、直前の発音からの時間差がわかればよい。この時間差は直前の音符の生成された局所テンポから算出できる。離鍵時刻は、現在の音符の拍長および、生成された局所テンポと演奏長比から算出する。そして、これに音高や生成された音の強さなどを付加し、その音符の演奏情報としてファイルに出力する。

5.2.3 実装

このシステムを計算機上に実装した。プログラムは全て C 言語で記述した。プログラム中ではマルチスレッドやフォークなどは行っていない。このプログラムを、DEC Alpha (OSF/1) 及び Sun UltraSparc (Solaris) の上でコンパイルし、それぞれバイナリを作成した。コンパイルはそれぞれ OS に付属している純正のコンパイラを使用し、最適化レベルも正常に動作する範囲で最大の最適化を行なうようにした。なお、次節で述べる実験は主に DEC Alpha 上で行なった。

5.3 実験

Kagurame Phase-I を使用して演奏データの生成を行なった。本節では、その演奏表情の生成実験について説明する。

5.3.1 準備

実験曲 まず、実験対象として Czerny の「8小節の練習曲」から 4 分の 4 拍子及び 4 分の 2 拍子の比較的良好に似ている楽曲 6 曲を選択した。実験に利用した楽曲を表 5.6 に示す。いずれの楽曲も 16 ないしは 32 分音符で構成されており、長さは全て 8 小節である。これらの楽曲は楽譜上では和音を含んでいるが、実験ではいずれも主旋律のみを対象とした。主旋律は単旋律である。

楽譜情報と楽曲構成情報 表 5.6 に示す楽曲の楽譜情報と楽曲構成情報は人手で作成した。ファイルの書式は 5.2.1 項で述べた通りである。楽曲構成情報はいずれの楽曲も二分木であるものとした。全ての旋律

番号	拍子	調	小節数	音符の総数
1	4分の4	C	8	121
12	4分の4	Bb	8	113
15	4分の2	G	8	113
24	4分の4	Bb	8	115
37	4分の4	A	8	121
65	4分の2	Bb	8	121

表 5.6: 実験曲

断片は原則として長さが同じ子を2つ持っている。ただし、音符の数や長さが合わない場合はこの限りではない。実験で用いた楽曲構成の一例を図 5.10 に示す。

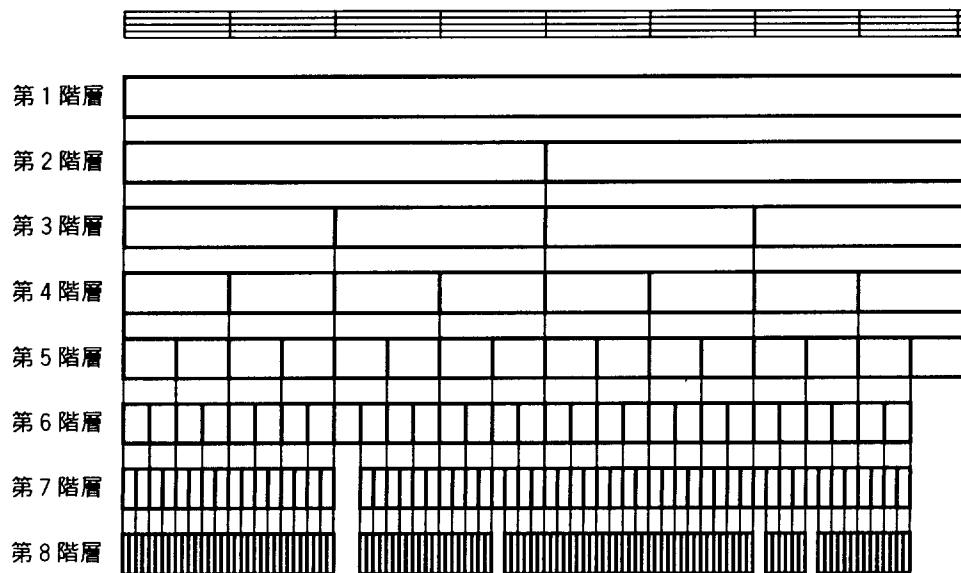


図 5.10: 実験に用いた楽曲構成の一例 (8小節の練習曲の24番)

演奏データ それぞれの楽曲について、以下の二種類の演奏状況の演奏データ、計 12 本を用意した。

- 古典派風
- ロマン派風

この演奏状況は演奏のスタイルに関するものである。古典派風の演奏はテンポを楽譜通りに正確に演奏し、2拍や4拍の周期で音の強弱を繰り返す点が特徴である。それに対して、ロマン派風の演奏は、楽曲の盛り上がりなどにあわせてテンポや音の強さを大きく変化させる傾向にある。

。 いずれの演奏表情も特徴的で、その相違を知っている人であれば容易に区別が可能である。また、ここで述べたような演奏表情の与え方の指針は演奏法として確立されたものであるため、この指針に基づいてある程度客観的に演奏表情の妥当性を判断することが可能である。これらの演奏表情は長い範囲で生じる

テンポの変化などを含んでいるため、適切な演奏表情を生成するためには、拍や小節などの短い範囲だけではなく、さらに長い範囲での旋律の流れなどを考慮する必要がある。

演奏は音楽大学で演奏法について学んだ演奏教育者1名によるものである。演奏ごとに、演奏曲の番号と上記の演奏スタイルのいずれかを指定して演奏者に演奏させた。演奏後に演奏データと楽譜とを比較し、弾き損じなどがなかったかを確認した。楽譜と合致しない部分があった場合は、演奏者にその旨を伝え、同じ曲を再度収録し直した。演奏には Roland の電子ピアノ HP-3000S を使用した。これを、PC/AT 互換機上の Windows98 で動作するシーケンスソフト Cakewalk8 で記録し SMF に保存した (図 5.11)。全ての演奏データは同じ日に演奏したものである。

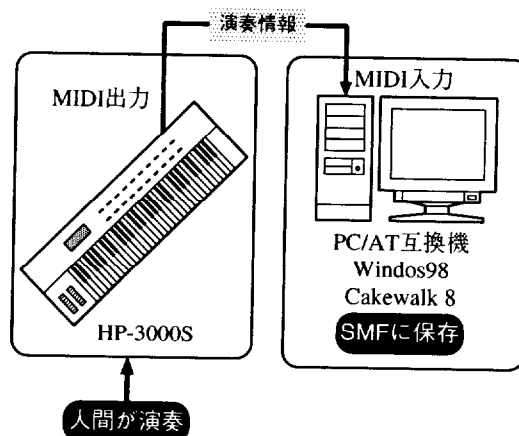


図 5.11: 演奏データの収集環境

5.3.2 生成

生成実験は、オープンテストとクローズドテストの二種類の手法で行なった。

クローズドテスト まず、クローズドテストの手順について概説する。クローズドテストでは、あらかじめ用意した 6 曲の 2 種類の演奏状況の演奏データ 12 本全てを、そのまま演奏データ集として利用した。そして、表 5.6 の各曲について、kagurame Phase-I を用いて表 5.7 に示す二種類の演奏表情を生成した。よって、生成した演奏データは合計で 12 本である。

	ロマン派風	古典派風
演奏状況 1	1	0
演奏状況 2	0	1

表 5.7: クローズドテストで用いた演奏状況設定

この実験では、演奏データ集の中に演奏対象曲のデータが含まれている。従って、理想的には人間の演奏と全く同じ演奏データが生成できるはずである。

オープンテスト オープンテストではクロスバリデーションにより、演奏対象曲の演奏データを演奏データ集に含まないようにして実験を行なった。オープンテストは、以下の手順で行なった。

1. 実験曲 表 5.6 の中から演奏対象曲を 1 曲選択
2. 演奏対象曲以外の実験曲 5 曲の演奏データ 10 本を演奏データ集とする
3. Kagurame Phase-I で演奏対象曲の演奏データを生成
4. 実験曲 6 曲全てに対して上記の手順を実行

それぞれの楽曲について、表 5.8 に示す 5 種類の演奏状況設定を与えた演奏データの生成を行なった。従って、オープンテストで生成された演奏データは合計で 30 本である。

	ロマン派風	古典派風
演奏状況 1	1	0
演奏状況 2	0.7	0.3
演奏状況 3	0.5	0.5
演奏状況 4	0.3	0.7
演奏状況 5	0	1

表 5.8: オープンテストで用いた演奏状況設定

通常の使用状況では、演奏対象曲の演奏データが演奏データ集に含まれている可能性は低いため、この実験の結果が一般的な状況での Kagurame Phase-I の能力に近いと考えられる。

録音 オープンテストで生成された SMF ファイルの中から、主に演奏状況設定を「ロマン派風:1, 古典派風:0」としたものと「ロマン派風:0, 古典派風:1」としたものを再生し、カセットテープに録音した。再生は PC/AT 互換機上の Windows98 で SMF ファイル再生プログラムを使用して行なった。PC/AT 互換機から出力された MIDI 信号を YAMAHA の Clavinova CLP-760 に入力して、その演奏音をカセットテープに録音した。録音環境を図 5.12 に示す。この録音テープは聴取実験で使用した。なお聴取実験については 5.4 節で述べる。

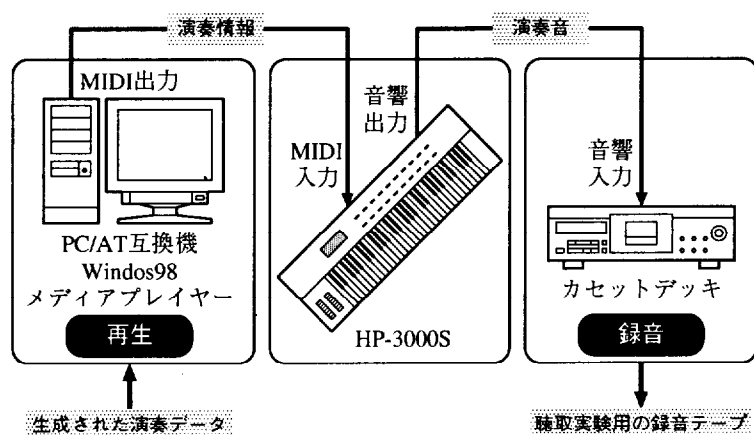


図 5.12: 演奏データの録音環境

5.4 評価

5.3 節で述べた手続きで Kagurame Phase-I に生成させた演奏データと、演奏データ集として利用した人間の演奏を比較し、Kagurame Phase-I の生成した演奏について評価を行なった。評価は人間による聴取と数値的な比較で行なった。

聴取による評価では、オープンテストで生成された演奏データを音楽大学でピアノ演奏について学んだ演奏経験者および音楽教育者 2 名に聴かせ、演奏技術や演奏法などの妥当性の観点から演奏について評論してもらった。聴取者の内 1 名は演奏データ集のデータの演奏者である。数値による比較では、主に Kagurame Phase-I が生成した演奏データと人間の演奏データとの差異や、Kagurame Phase-I が生成した二種類の演奏状況のデータの間の差異を比較した。

5.4.1 再現能力

まず最初に、クローズドテストで生成されたデータから、Kagurame Phase-I の演奏表情の再現能力を評価した。クローズドテストでは、演奏対象曲の演奏データが既に演奏データ集の中に含まれているため、システムにとって有利な実験である。

ここでは、人間の演奏にどれだけ近い演奏が再現されているかについて、数値的な比較から評価した。実験に利用した楽曲ごとに、演奏データ集に利用した人間による演奏のデータと Kagurame Phase-I が生成した演奏データとの差異を計算した。「ロマン派風:1, 古典派風:0」として生成したデータとロマン派風の人間の演奏、「ロマン派風:0, 古典派風:1」として生成したデータと古典派風の人間の演奏をそれぞれ比較した。人間の演奏データ P_h とシステムの演奏データ P_s の差異 $d(P_h, P_s)$ の計算式を式 (5.11) に示す。

$$d(P_h, P_s) = \frac{1}{|S|} \cdot \sum_{n \in S} \left| \frac{\exp(P_h, n)}{\sum_{n \in S} \exp(P_h, n)} - \frac{\exp(P_s, n)}{\sum_{n \in S} \exp(P_s, n)} \right| \quad (5.11)$$

S は演奏対象曲に含まれる音符の集合、 $|S|$ は演奏対象曲の音符の総数、 n は演奏対象曲の任意の音符を表す。 $\exp(P, n)$ は演奏データ P の音符 n の局所的な演奏表情値（テンポ、音長、ヴェロシティのいずれか）である。演奏の差異は、音符ごとの局所的な演奏表情値を楽曲全体の演奏表情値の平均で割った数値の差の絶対値の平均である。演奏の差異 $d(P_h, P_s)$ が小さいほど、人間の演奏と Kagurame Phase-I の出力が近いことを示す。二つの演奏が全く同じであれば演奏の差異は 0 になる。

人間の演奏とシステムの演奏との差異を表 5.9 に示す。また、参考値としてシステムの生成したロマン派風の演奏と古典派風の演奏との間の差異を表 5.10 に示す。表 5.9 から明らかなように人間の演奏と生成された演奏の間の差異は殆んどない。一番大きく異なっているロマン派風の 1 番でも、その差異は 0.01 % に満たない。スタイルの異なる演奏間の差異が標準的な値であるとすれば、差異の標準値は軒並 10 % を越える値である。これに比べると、人間の演奏とシステムとの演奏はいずれも差がないに等しい。このことは図 5.13 のグラフからも明らかである。このグラフは人間のロマン派風の演奏と、演奏状況を「ロマン派風:1, 古典派風:0」として生成した演奏のテンポを比較したものである。演奏曲は 8 小節の練習曲の 1 番である。横軸が小節、縦軸は演奏表情の比率を表している。このグラフでは二つの線が完全に重なっており、その差異は確認できない。

規則に基づく手法では、規則を学習した時点で個々の演奏データの演奏表情の情報が失われてしまうため、学習などに利用した演奏データに近いデータが再現される可能性は稀である。それに対して、この実験で Kagurame Phase-I が生成した演奏データは、人間の演奏データとはほぼ同じである。

4.1 節や 4.2 節で述べたように、事例に基づく演奏表情生成手法では、子細な類似性の低い事例を参考事例として利用したり、演奏表情を比率値に分解するなどの複雑な操作を行なっている。これらは、生成される演奏表情の品質を低下させる要因ともなり得るが、この評価の結果から、これらの演奏表情処理技術

演奏状況	演奏曲	テンポ	音の強さ	演奏長比
ロマン派風	1	0.0009	0.0009	0.0008
	12	0.0005	0.0003	0.0006
	15	0.0004	0.0003	0.0004
	24	0.0005	0.0005	0.0004
	37	0.0003	0.0002	0.0004
	65	0.0004	0.0003	0.0003
古典派風	1	0.0005	0.0005	0.0006
	12	0.0003	0.0003	0.0005
	15	0.0004	0.0002	0.0004
	24	0.0003	0.0004	0.0005
	37	0.0003	0.0003	0.0003
	65	0.0003	0.0002	0.0004
平均値		0.0004	0.0004	0.0005

表 5.9: クローズドテストでの人間の演奏とシステムの演奏の差異

演奏曲	テンポ	音の強さ	演奏長比
1	0.1483	0.1071	0.0895
12	0.1315	0.0975	0.0985
15	0.1846	0.1181	0.1378
24	0.1614	0.0858	0.1566
37	0.1572	0.0958	0.1037
65	0.2007	0.1541	0.1484
平均値	0.1640	0.1097	0.1224

表 5.10: システムがクローズドテストで生成した演奏状況の異なる演奏データの差異

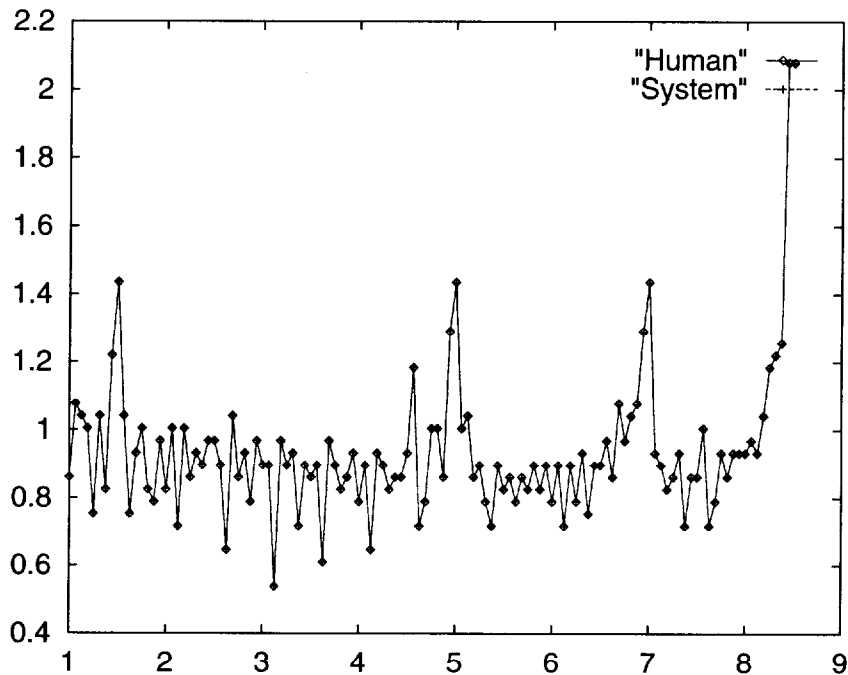


図 5.13: クローズドテストで生成した演奏と人間の演奏との比較 (Czerny の 8 小節の練習曲の 1 番のロマン派風の演奏のテンポ)

の導入が、演奏データの再現能力に殆んど影響を与えないことが確認できた。

この結果から一概に事例に基づく手法の基本能力を断言することはできないが、少なくとも有効な演奏データがあった場合の演奏表情生成能力が高いことは明らかである。

5.4.2 演奏技術

本項以降は、オープンテストによって生成したデータに対する評価について述べる。オープンテストでは、演奏データ集に演奏対象曲のデータが含まれていないため、この評価は一般的な状況でのシステムの能力を評価したものである。

まず、Kagurame Phase-I が生成した演奏データの演奏技術に関する評価について述べる。生成された演奏データの演奏表情は概ね人間の演奏のように聴こえる自然なものであった。演奏データ集の中には弾き損じにより極端に弱くなってしまった音が見られたが、生成されたデータでは極端に弱い音は見られなかった。複数の事例を参考にしてしているため、単独の事例にしか見られないこういった一種の演奏ミスがある程度吸収されたためであると考えられる。

生成された演奏表情の変化は人間の演奏によく似た傾向を示すところが多かった。特に演奏表情の変化が顕著な部分で人間の演奏に近くなることが多い。また、その多くは一般に演奏表情の与え方の特徴として良く知られているものであった。これは、一般的な演奏表情であるため、このような演奏表情を持った事例が数多くあったのではないと思われる。

次に、人間の演奏との相違について述べる。演奏データの一例として、人間の演奏とシステムが生成した演奏のテンポ変化、音の強さ、演奏長比の比較をそれぞれ図 5.14、図 5.15、図 5.16 に示す。

これらはツェルニーの八小節の練習曲の 37 番の演奏データである。システムの演奏は演奏状況を「ロマン派風:1, 古典派風:0」としたもので、人間の演奏は「ロマン派風」の演奏データである。図 5.14 の縦軸は

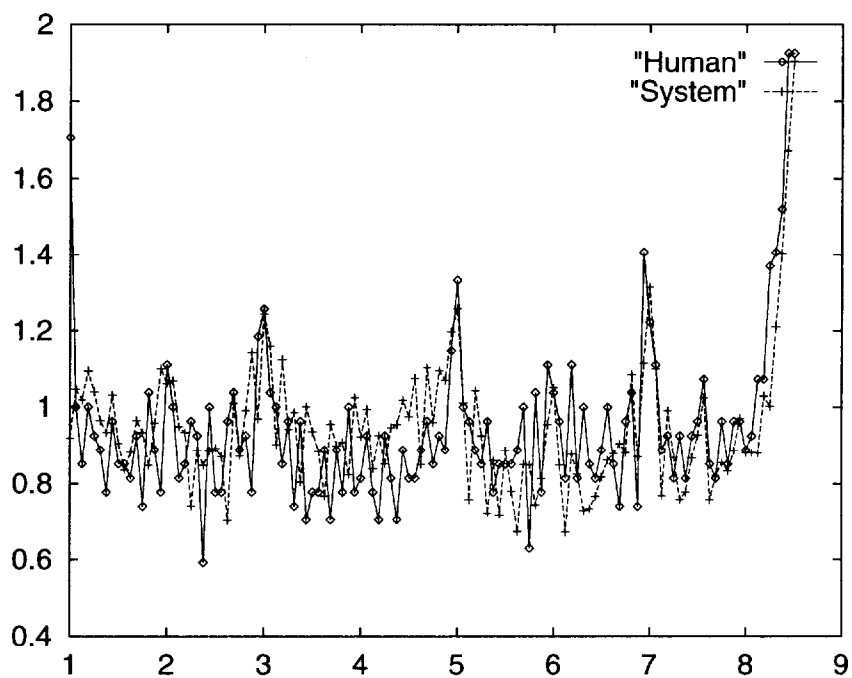


図 5.14: オープンテストで生成した演奏と人間の演奏との比較 (8 小節の練習曲の 37 番のロマン派風の演奏のテンポ)

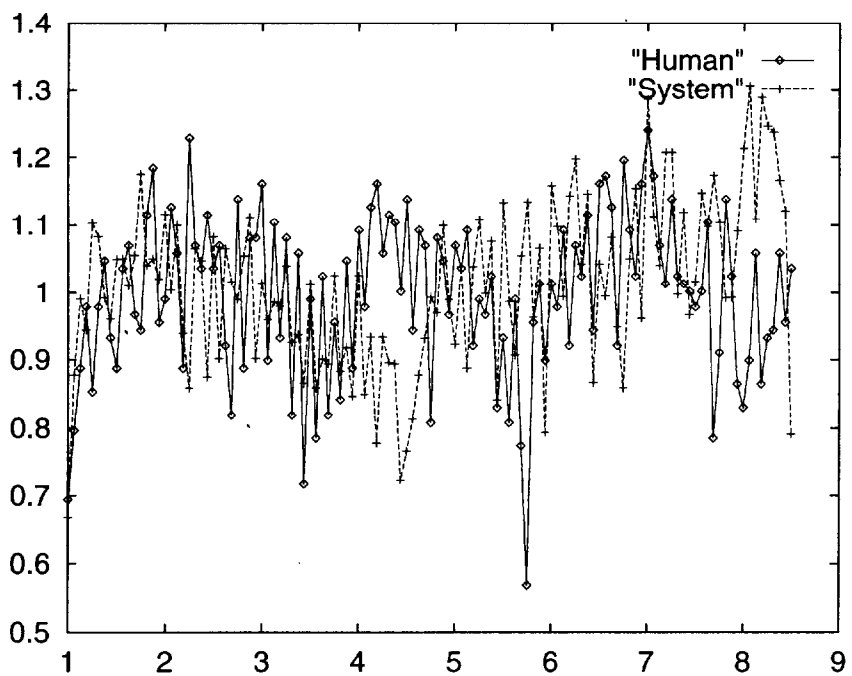


図 5.15: オープンテストで生成した演奏と人間の演奏との比較 (8 小節の練習曲の 37 番のロマン派風の演奏の音の強)

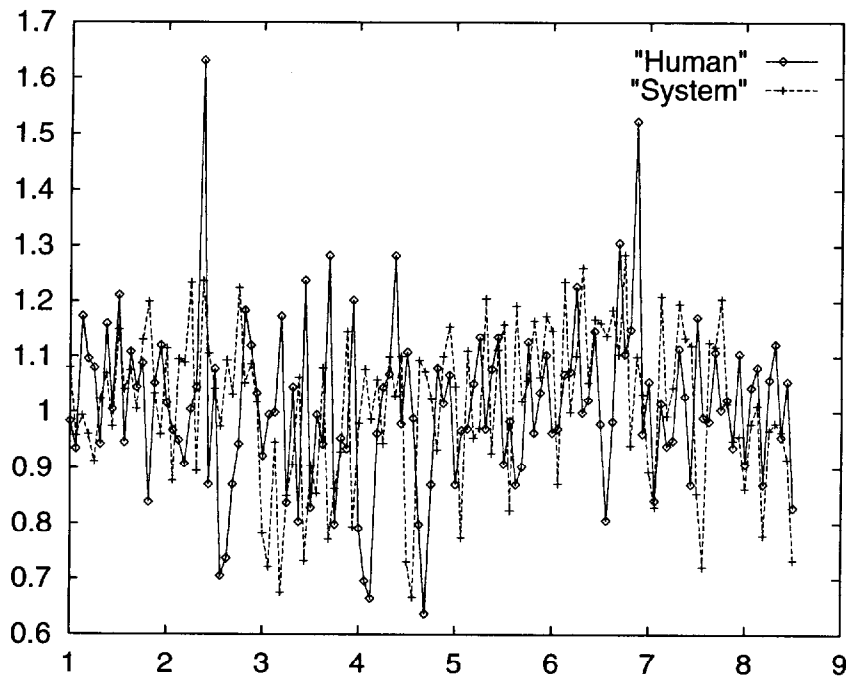


図 5.16: オープンテストで生成した演奏と人間の演奏との比較 (8小節の練習曲の37番のロマン派風の演奏の演奏長比)

演奏データ全体のテンポの平均値を音符単位の局所テンポで割った数値である。数値が大きいほどテンポが遅いことを表す。図 5.15 の縦軸は音符ごとの音の強さの値を演奏データ全体の音の強さの平均値で割ったもので、数値が大きいほど音が大きいことを表す。図 5.16 の縦軸は音符ごとの演奏長比を演奏データ全体の演奏長比の平均値で割ったもので、数値が小さい部分がスタッカートに近い演奏である。横軸のいずれも小節である。テンポのグラフからは 3 小節目の冒頭や 5 小節目の冒頭、7 小節目の冒頭、終末部分など、テンポが極大値と極小値をとっている場所が一致しており、全体的なテンポの推移も類似している。また、音の強さのグラフも前半部分や 6 ~ 7 小節目などの変化の傾向が類似している。演奏長比も全体的な変化の傾向や数値の極大値と極小値の周期性などが類似している。このような傾向は他の楽曲の演奏データでも見られる。特に、テンポに関してはいずれの演奏データもこのような類似した結果が得られた。

一方で、人間の演奏とシステムの演奏が明らかに異なる部分も見られた。その一例として 1 番の演奏データの音の強さの推移を図 5.17 に示す。このグラフでは 1 ~ 3 小節目や 8 小節目の推移は類似しているが、4 小節目の後半から 7 小節目にかけての全体的な推移が大きく異なっている。このような推移の大きな相違は、音の強さで多く見られた。人間の聴取による評価では、このような相違の多くは間違いと言うほどのものではないという意見が得られた。今回は演奏状況を「ロマン派風」と「古典派風」の 2 つの素性で記述したが、同じロマン派風に分類される演奏表情の中にも様々な違いがあり、異なる部分の多くはこういった演奏状況に記述されていない細かな演奏状況の相違によって生じているのではないかと考えられる。より人間の演奏に近い結果を得るためには演奏状況を現在より詳細に記述する必要があると考えられる。

演奏長比に関しては、もともと大きな変化が少なく、分布の変化も楽曲全体に渡ってほぼ均一であるため、極端な相違はあまり見られず、全般的に人間の演奏と同じような傾向を示していた。

人間の演奏とシステムの演奏とどちらがよいかという質問に対しては、人間の演奏の方がよいという意見が多かった。一方で生成された演奏が単旋律なので評価が困難であるという意見もあった。

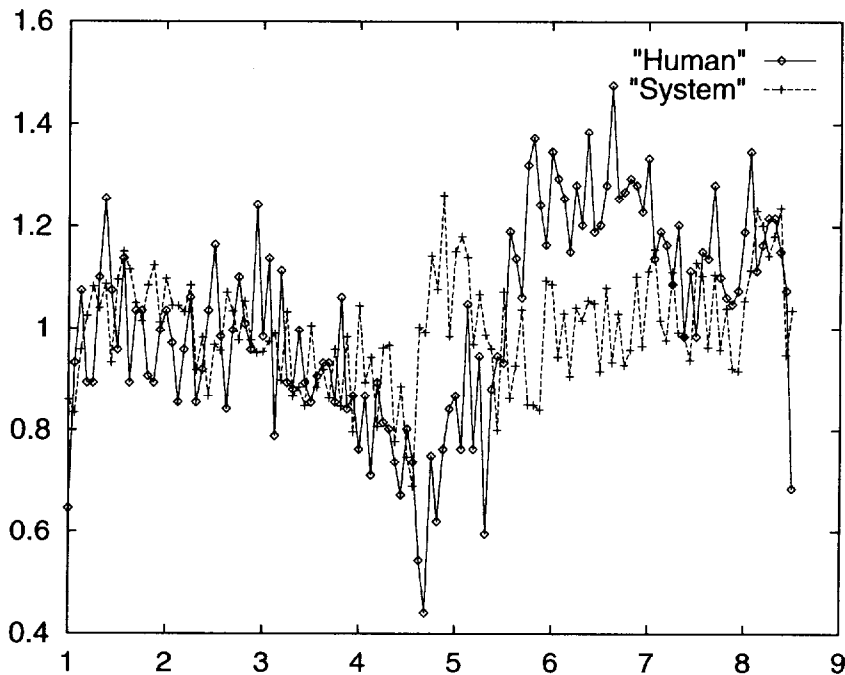


図 5.17: オープンテストで生成した演奏と人間の演奏との比較 (8 小節の練習曲の 1 番のロマン派風の演奏の音の強さ)

5.4.3 演奏状況による演奏表情の相違

次に演奏状況に応じた弾き分けに関する評価について述べる。生成された二種類の演奏はいずれの楽曲も明らかに異なるものであった。

図 5.18 にシステムが生成した「ロマン派風:1, 古典派風:0」と「ロマン派風:0, 古典派風:1」の演奏データのテンポの比較を示す。演奏曲は 37 番である。ロマン派風は古典派風に比べてテンポの変化の範囲が広がっており、この点は一般的に知られているロマン派風の演奏の特徴とも一致している。この傾向は全ての演奏データで見られた。

また、生成された二種類の演奏状況の演奏データを聴取者によって「ロマン派風」か「古典派風」かの判定をさせたところ、判定に使用した全ての楽曲について正解が得られた。これらから、少なくとも今回与えた二種類の演奏状況の弾き分けは適切に行なわれていると評価できる。

次に、演奏状況をロマン派風と古典派風の間設定した演奏データを図 5.19 示す。これは、37 番の演奏データの音の強さである。このグラフは、表 5.8 に示す 5 種類の演奏状況の演奏データをプロットしたものである。演奏状況を中間的な値に設定した演奏データは、それぞれ素性の程度量の比率に応じた中間的な演奏データになっている。

5.4.4 計算時間

5.3 節のオープンテストで演奏状況 1 および、演奏状況 5 の演奏データの生成に要した時間を表 5.11 に示す。この数値は、AlphaServer 4100 (Alpha 21164 400MHz 4 CPU, Memory 3G byte) 上で Kagurame Phase-I を実行し、演奏表情の生成を行なった際の CPU 時間を秒単位で表したものである。

一曲の生成時間は平均で 1 分強である。また、生成に要する時間は音符数にほぼ比例しており、演奏状

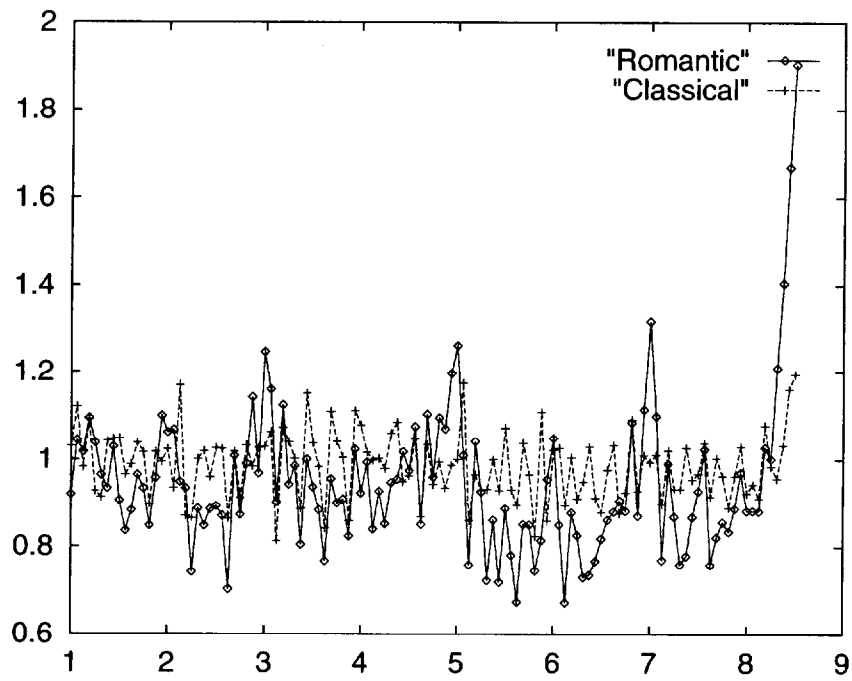


図 5.18: オープンテストで生成したロマン派風と古典派風の演奏の比較 (8小節の練習曲の37番の演奏のテンポ)

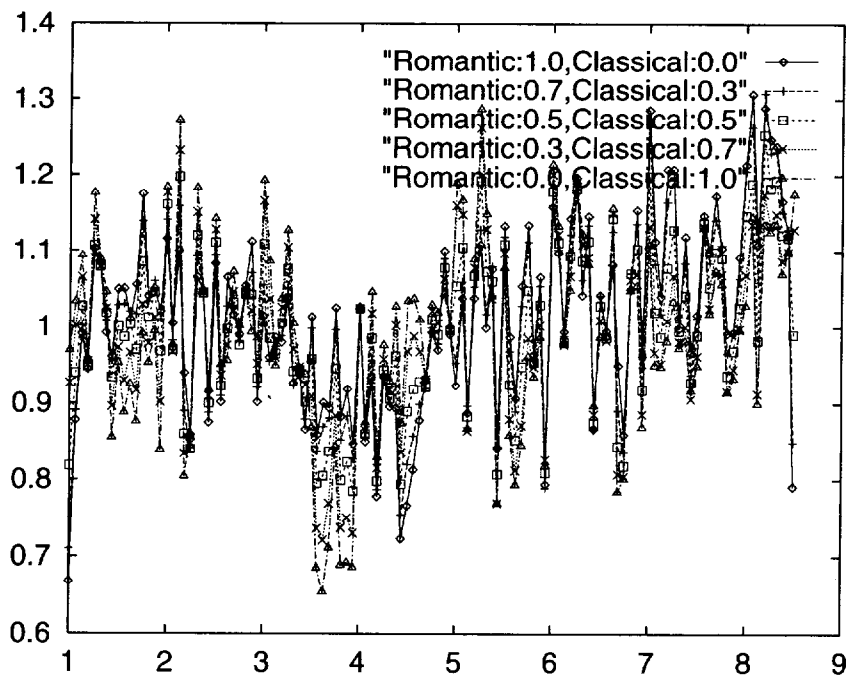


図 5.19: 演奏状況が中間的な演奏データの比較 (8小節の練習曲の37番の演奏の音の強さ)

演奏曲	音符数	ロマン派風	古典派風
1	121	70.6 秒	70.6 秒
12	113	67.8 秒	67.9 秒
15	113	66.9 秒	67.4 秒
24	115	68.3 秒	68.3 秒
37	121	71.3 秒	71.3 秒
65	121	74.8 秒	74.8 秒
平均		70.0 秒	

表 5.11: 演奏表情の生成に要した時間 (秒)

況には殆んど依存していない。

演奏表情生成の各段階に要した時間の割合を表 5.12 に示す。オープンテストで演奏状況 1 および、演奏状況 5 の演奏データを生成した際の平均値である。この表からも明らかなように、生成の殆んど時間は参考事例の検索に費やされている。Kagurame Phase-I では、参考事例の検索の際に、演奏対象曲の旋律断片を演奏データ集の全ての旋律断片と比較しているため、この部分の計算コストが高く、これが計算時間にも反映されている。

処理	時間割合
データの読み込み	0.22 %
参考事例の検索	99.75 %
重要度の評価と演奏表情の分析	0.01 %
演奏表情の合成	0.00 %
演奏表情の適用	0.00 %
演奏データの出力	0.02 %

表 5.12: 演奏表情生成の各段階での所要時間の割合

第6章 自動伴奏システムへの応用

自動伴奏システムは、独奏部分を演奏する人間と協調して合奏を行なうことを目的としたシステムである。自動伴奏システムに事例に基づく演奏表情生成手法を導入することにより、従来から伴奏システムで課題となっていた、伴奏の演奏表情の作成や演奏者の演奏傾向の学習といった問題を解決することが可能である。本章では、伴奏システムの持つ課題と、それに対して事例に基づく演奏表情生成が如何にして機能するかについて述べる。

6.1 自動伴奏システムについて

6.1.1 概要と先行研究

自動伴奏システムとは、人間（独奏者）の演奏に合わせてリアルタイムに伴奏を演奏するシステムである [27, 32]。本節では、まず自動伴奏システムについて概説する。

自動伴奏システムの基本的な発想はミュージックマイナスイオン CD やカラオケに見ることができる。ミュージックマイナスイオン CD とはコンチェルトなどの楽曲の、独奏のパートを除いた残りのオーケストラの演奏を録音した CD のことである。これは、独奏者がコンチェルトの練習などに利用することを目的としたものである。ポピュラー音楽のシングル盤などに収録されているカラオケ演奏（伴奏のみの録音）もこれに類するものである。

カラオケは主に、LD や CD に録音された伴奏の演奏や、あるいはコンピュータ上に記録された伴奏の演奏情報を再生し、それと同時に歌詞を画面に表示するシステムである。

これらのシステムの利点は、独奏者単独でも合奏を行なうことが可能である点である。一般に、合奏を行なうためには複数の演奏者を必要とするが、オーケストラのような大編成の楽曲の場合、演奏者を集めることはそう容易ではない。そこで、自分以外の演奏者を CD やカラオケ装置に肩代りしてもらうことで、独奏者単独での合奏が可能となる。

しかし、これらのシステムは、機械が一方的に行なう演奏に対して、人間が合わせて演奏しないと合奏が成立しないという欠点がある。本論文で何度も述べているように人間の演奏には演奏表情が含まれている。これは合奏においても例外ではなく、演奏者は各々が演奏をより自然なものにするために様々な演奏表情を付与して演奏を行なう。しかし、合奏の場合は各々が好き勝手に演奏をしたのでは、合奏全体が成り立たなくなってしまうため、演奏者同士がお互いの演奏を聞いてそれに合わせつつ共通の演奏表情を与えるように勤めている。こういった、演奏者同士の駆け引きの中に合奏の醍醐味ともいえるものがあるともいわれている。

カラオケやミュージックマイナスイオン演奏は、その構成は合奏に近いものであるが、システム側では人間の演奏を一切考慮せず一方的に演奏するのみである。そこには今述べたような駆け引きは一切なく、また演奏表情も伴奏システムに与えられるがままである。カラオケシステムの中には、伴奏のテンポや音の強さを修正する機能がついているものもあるが、これらも人間がテンポの変更を明示的に指定しなければならないため、結局のところ演奏時は完全に人間がシステムに合わせて演奏することになる。

カラオケやミュージックマイナスイオン演奏のこれらの問題を解決したものが自動伴奏システムである。自動伴奏システムは、独奏者の演奏を「聴く」能力を持っており、独奏者の演奏に合わせて伴奏のテンポ

や音の強さを修正することが可能である。

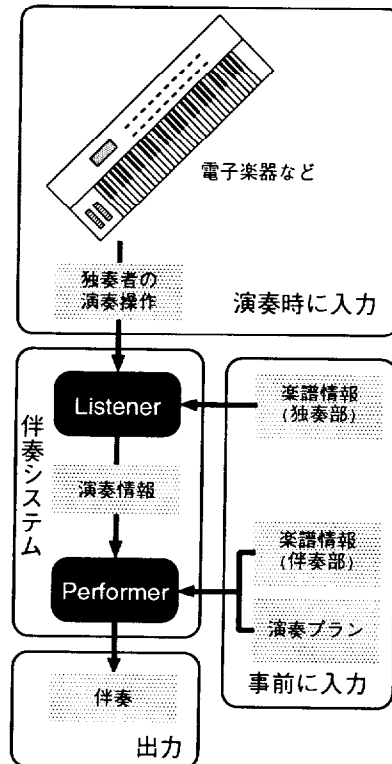


図 6.1: 自動伴奏システムの標準的な構成

図 6.1 に伴奏システムの標準的な構成を示す。これは、電子楽器を用いた自動伴奏システムの標準構成である。音響信号を入力として利用する場合は、この他に、前処理としてピッチ抽出やアタック検出を行なう部分が含まれる。

自動伴奏システムは、あらかじめ人間が担当するパートとシステムが担当する伴奏パートの楽譜をそれぞれ与えておく。また、伴奏部分の演奏表情をいかにして与えるかの指針である演奏プランも同様にあらかじめ用意する。演奏プランについては 6.1.2 項 で説明する。

標準的な自動伴奏システムは Listener と Performer という二つの構成要素を持っている。

- Listener

Listener は独奏者の演奏を分析する部分である。合奏時には、まず、Listener が独奏者の演奏操作をリアルタイムに受信する。独奏者の演奏操作の入力方法は、伴奏システムによって異なるが、電子楽器からの出力信号 (MIDI) を直接利用する方法と、演奏音を信号処理しその結果を利用する方法が一般的である。Listener では、入力された演奏操作と独奏部分の楽譜との対応づけを行ない、独奏者が現在楽譜上のどこを演奏しているかを推定する。また、演奏者の演奏には、弾き間違いや弾き損じが含まれている可能性があるため、単純に楽譜上の次の音符と比較するだけではなく、長めの範囲内にある音符と比較し、弾き間違いや弾き損じに対応する必要がある。さらに、Listener では、楽譜情報、演奏操作の種別、そして演奏操作が行なわれた時刻から、テンポなどの独奏者の演奏表情情報の計算も行なう。この出力は、独奏者の演奏情報として Performer に送られる。

- Performer

Performer は伴奏部分の演奏を行なう部分である。Performer では、まず、独奏者の演奏情報を受けて、伴奏部分のテンポや音の強さなどの演奏表情をどうするかを決定する。演奏表情の決定の際には、それまでの独奏者の演奏操作の履歴の分析や、それ以降に行なわれるであろう独奏者の演奏操作の予測を行なう。その結果を総合的に判断し、あらかじめ用意した演奏プランを独奏者の演奏により適合するように修正することで伴奏部分の演奏表情や演奏タイミングを決定する。そして、その結果にしたがって伴奏部分の楽譜をリアルタイムに逐次的に演奏し出力する。この伴奏部分の演奏が独奏者の演奏と協調することで合奏が成立する。

自動伴奏システムに関する研究は、Dannenberg らの伴奏システム [7, 10], Vercoe らのシステム [72], 堀内らの自主性を持つ伴奏システム [25, 30] などが報告されている。なお、ジャズの即興システム [3, 18, 21, 50] のように、人間の演奏を受けて伴奏部分を作曲もしくは編曲し、その上で演奏を行なうシステムを自動伴奏システムに含めることもあるが、ここではこれらの創作活動を伴うものは除外する。

自動伴奏システムで最も重要なことは、いかにして伴奏の演奏を人間の演奏に合わせるかということである。そのためには、Listener で行なっている独奏の演奏と楽譜との照合、Performer で行なっている独奏者の演奏の分析や予測、そしてそれに基づく演奏表情の決定が重要である。先行研究でもこの点に重点をおいて研究を行なっている。

ここでは、それぞれのポイントごとに先行研究で提案されている技術を概観する。

- 独奏の演奏と楽譜との照合

独奏者の演奏は、弾き間違いや弾き損じなどの演奏ミスによって、意図せずに楽譜と異なる演奏を行なってしまふことがある。自動伴奏システムはこれら独奏者の演奏ミスに対応できなければならない。そのためには、独奏の演奏と楽譜との照合が演奏ミスに対して頑健である必要がある。そこで、自動伴奏システムでは DP マッチングなどの、演奏ミスに頑健な楽譜照合アルゴリズムを利用してゐる。また、演奏ミス時に発生する楽譜追従の失敗からの復帰を早めるため、音程情報だけではなく時間情報や演奏ミスの傾向分析などを利用した手法 [7] も提案されている。

- 独奏者の演奏の予測

伴奏の演奏表情の中で最も難しいのがテンポの制御である。テンポのずれは、独奏と伴奏の発音のずれとなって顕著に現れるため、音の強さなどの他の演奏表情に比べて、制御の失敗の影響が大きい。伴奏部分のテンポのスケジューリングを行なうためには、独奏者のテンポ変化を予測する必要がある。先行研究では、独奏者の演奏の進行速度に従って進む仮想的な時計を仮定し、この時計を利用して将来の独奏のタイミングを予測している。仮想時計の速度は直前の独奏者の演奏履歴から随時調整する。

- 演奏表情の決定

伴奏部分の演奏表情は、伴奏部分の演奏表情の指針である演奏プランに基づいて与えられる。しかし、演奏プラン通りに伴奏を演奏したのではカラオケや CD と何らかわらない一方的な演奏になってしまうため、演奏表情の決定の際に、演奏プランを独奏の演奏に適合するように修正する必要がある。演奏プランの修正に当たっては、完全に独奏に合致するように修正してしまうと、伴奏部分の主体性がなくなってしまう、結果として伴奏が逆に不自然になってしまうことがある。そのため、演奏の主導権という概念を導入し、状況によっては独奏をあまり考慮せず演奏プランに近い伴奏を行なう手法 [30] などが提案されている。

- この他に、複数の独奏者との協調を行なう自動伴奏システム [26, 31] や、自動伴奏システムのための音響認識技術の研究 [19, 35] が行なわれている。

6.1.2 演奏プラン

演奏システムと同様に、自動伴奏システムも自然な合奏を行なうためには演奏表情の知識が必要である。第5章で述べた演奏システムのような単独での場合、演奏表情は演奏者、つまり演奏システムの意思のみで決定しても問題がなかった。しかし、自動伴奏システムのような複数の演奏者が協調するシステムの場合、他の演奏者の演奏表情を考慮した上で演奏時にリアルタイムに演奏表情を決定する必要があるため、事前に演奏表情を決定することはできない。かといって、演奏時に全ての演奏表情を生成した上で決定するのも困難である。そこで、自動伴奏システムでは、伴奏部分の雛型的な演奏表情をあらかじめ作成して付与している。これが演奏プランである。伴奏システムの演奏表情は、基本的に演奏プランを独奏者の演奏にうまく合うように修正することによって決定される。従って、演奏プランは伴奏システムの演奏表情に対して大きな影響力を持った情報である(図6.2)。

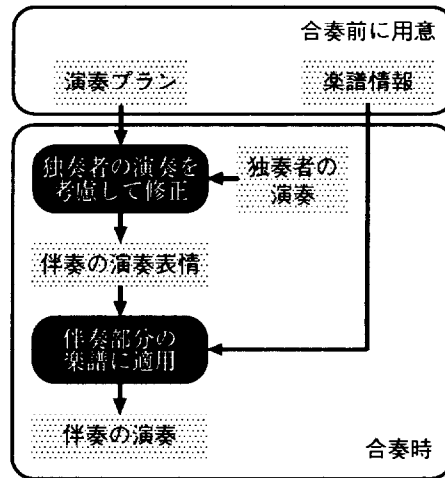


図 6.2: 演奏プランと演奏表情との関係

演奏プランという概念は、自動伴奏システム固有のものではない。人間の演奏者は、合奏時にどのような演奏表情を与えるかの予定や考えをあらかじめ持っており、それに基づいて合奏を行なっていることから、人間の演奏者も演奏プランに類する知識を持っていると考えられている。勿論、実際の合奏時には、他の演奏者の演奏に演奏を合わせる必要があるため、独奏者の場合も演奏プランが直接演奏表情になるわけではない。

自動伴奏システムの演奏プランは、演奏曲の楽譜に基づいて、標準的であると思われるものを人手で与えている。演奏プランの細かい部分、特に具体的な数値には入力者の主観的な判断が必要である。この作業にはある程度の演奏経験と熟練が必要であり、自動化が望まれる。演奏プランは演奏表情に近い情報であるため、演奏表情の自動生成技術を応用することで演奏プランの生成も可能であると考えられる。しかしながら、演奏プランの自動生成に関しては殆んど研究が行なわれていない。

また、独奏者の演奏プランは演奏表情と同様に演奏者によって異なるため、全ての演奏者に対して同じ演奏プランを利用するのは望ましくない。事実、自動伴奏システムで実際に合奏を行なってみると、システムの演奏プランと独奏者の演奏プランが大きく違っている場合に、合奏時の演奏がうまく噛み合わないという問題が発生する。この問題を回避するためには、独奏者の趣向に合わせた演奏プランの作成や修正を行なう必要がある。この演奏プランの修正を行なう手法の一つに次節で説明するリハーサルがある。

6.1.3 自動伴奏システムとリハーサル

演奏者間の演奏プランの相違は、コンピュータと人間との間に限って起きることではなく、人間同士でアンサンブルを行なう場合にも頻繁に発生することである。むしろ、初対面の二人がアンサンブルを行なってお互いに納得のいく演奏ができることはまずない。そこで、人間同士の合奏の場合、演奏者間の演奏プランの相違を吸収するために、あらかじめリハーサルを行なう。リハーサルでは、演奏者同士が試しに合奏を行なう。そして、その演奏から互いの演奏プランの相違を認識し、自分の演奏プランを相手の演奏プランに近付けるよう修正する。この作業を何度か繰り返すと、最終的に演奏者間の演奏プランの相違が小さくなり、結果として合奏がうまく合うようになる。

リハーサルは、自動伴奏システムの演奏プランと演奏者の演奏プランとの相違を解消するためにも有効な手段であると考えられており、自動伴奏システムへのリハーサルの導入の研究が行なわれている。

Vercoe は独奏者との合奏（リハーサル演奏）を何度も行ない、その記録を統計処理して独奏者固有の演奏の傾向を抽出するというリハーサルのモデルを提案している [73]。まず、リハーサル演奏ごとに記録された演奏データから、楽譜上の各音符ごとの独奏者の演奏表情の平均値及とその標準偏差を計算する。そして平均値を独奏者の望んでいる演奏表情とし、標準偏差を独奏者の演奏のばらつきとする。そして、独奏者の演奏表情の平均値にもとづいて、伴奏の各部の演奏プランを修正する。また、本番の演奏の際には、標準偏差を考慮して独奏者の演奏を認識することで、独奏者の意図しない演奏表情の揺らぎの影響を軽減するというものである。

Vercoe の手法は比較的簡単に独奏者の演奏プランに近い情報を取得することが可能である。しかしながら、有意な情報を収集するためには、リハーサル演奏を何度も行なわなければならない。あまりに多くの演奏を行なうことは独奏者に対してかなりの負担になるし、何度も繰り返し演奏することによって独奏者の演奏自体が変化してしまうことも考えられる。また、独奏者の演奏を分析するだけでは独奏者の演奏プランを的確に推定できない場合がある。独奏者はリハーサル時にも、無意識のうちにシステムの演奏にある程度合わせようとして演奏を行なうため、リハーサルにより独奏者の演奏プランとして得られるものは、実際にはシステムの演奏プランの影響を受けたものになる。例えば、伴奏のテンポが独奏者の意図しているテンポより極端に遅い場合、独奏者は最初は自分の意図するテンポではなく伴奏のテンポにあわせて演奏を開始し、自分の希望するテンポになるまで徐々にテンポを速めていくであろう。このリハーサル結果から得られる独奏者の演奏プランは、独奏者の意図した「もっと速いテンポでの演奏」ではなく「テンポを徐々に速くする」というものになってしまう。

堀内らは、リハーサル演奏に独奏者と自動伴奏システムの意見交換の枠組を導入したツーフェイズリハーサルという手法を提案している [28, 29]。人間同士のリハーサルを考えると、演奏者はただ単にお互いの演奏に合わせて演奏するだけでなく、その演奏によって得られた情報をもとに、相手の演奏プランに関する質問を行ったり、また逆に相手に自分の演奏プランを伝えて何らかの要求をしたりする。これらの意見交換は、演奏者同士の演奏プランをより近いものにする上で重要な役割を担っていると考えられる。そこで、従来のリハーサル演奏で行なっていた独奏者の演奏から演奏の癖や特徴を抽出する処理に加えて、独奏者との間で質問や要求を交換する手続きを新たに導入し、より人間同士のリハーサルに近いリハーサルモデルを実現したものがツーフェイズリハーサルである。ツーフェイズリハーサルでは、リハーサル演奏の手続きを「演奏フェイズ」、独奏者と自動伴奏システムとの意見交換を「コミュニケーションフェイズ」と呼んでいる。演奏フェイズとコミュニケーションフェイズの組をステージと呼び、このステージを何度か繰り返すことでリハーサルを行なう（図 6.3）。

演奏フェイズではその時点でシステムが保持している演奏プランを利用して独奏者との合奏を行なう。それと同時に独奏者の演奏と独奏部分の楽譜との比較を行ない、システムが予想していた演奏と実際の独奏者の演奏との相違を独奏者の演奏プランとして記録する。独奏者の希望によっては、演奏フェイズを曲の途中から演奏を開始したり途中で中断したりする。コミュニケーションフェイズでは、演奏フェイズで得

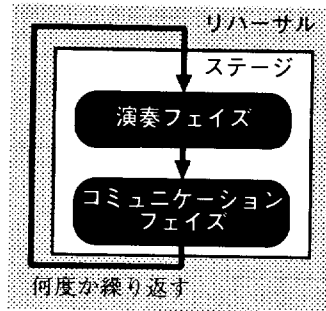


図 6.3: ツーフェイズリハーサル

られた独奏者の演奏プランに基づいて独奏者と対話を行なう。対話はシステムから独奏者への質問と独奏者からシステムへの要求の二種類からなる。独奏者への質問は、システムが抽出した独奏者の演奏プランの確信度が低い場合に、独奏者にシステムの解釈の是非を確認する。システムへの要求は、主にテンポや音量に対する独奏者の希望を直接入力する手段として利用する。

ツーフェイズリハーサルでは、統計的な処理は行なわず、演奏者の演奏に演奏者からの意見を加えたものを演奏者の演奏プランとしている。Vercoe の手法では、統計的に有意な情報を取得するためには何度も演奏を行なわなければならないが、ツーフェイズリハーサルではコミュニケーションフェイズの導入によって演奏回数が少なくても独奏者の的確な演奏プランが取得できる。

これらのリハーサル手法の問題点の一つとして、繰り返しなど楽曲中の類似した部分に対してリハーサルの効果がないことが挙げられる。通常の楽曲では、繰り返しや主題の再現など、同じ楽曲の中に類似した旋律が何度も出現することが多い。人間同士がリハーサルを行なう場合、類似した旋律のある一箇所についてリハーサルを行なえば、他の類似した部分については、リハーサルを行なわなくてもある程度合わせた合奏ができる。例えば、楽曲中の繰り返しがある部分で、繰り返しの一回目だけリハーサルを行なうと、繰り返しの二回目の演奏はすんなり合ってしまう。これは、これらの類似した部分に対して類似した演奏プランが適用されることが多いためである。従って、自動伴奏システムのリハーサルでも、楽曲中のある部分のリハーサルから得られた演奏プランを、他の類似した部分にも適用することで、リハーサルの効率化が可能である。しかしながら、従来のリハーサル手法は、楽曲の繰り返しや構造的な類似に関する情報を持っていないため、ある部分に対する独奏者の要求を曲中の類似した部分に適用することができない。従って、繰り返し部分では必ず繰り返しの回数分リハーサルをし、その都度同じ演奏や要求を行なわなければならない。

この問題を解決するためには、リハーサルの演奏記録をそのまま演奏プランとして利用するのではなく、演奏記録を楽曲の構成などに照らし合わせて分析し、その結果を再構成することで新たに演奏プランを構築する必要がある。

もう一つの問題点は、演奏者への適用である。人間の演奏者の場合、何度か合奏を行なったことがある人間同士であれば、一回目のリハーサルである程度協調した合奏ができてしまうことが多い。これは、過去の合奏の経験から相手がどのような演奏を行なうかが予測できるためである。つまり、演奏者は過去の合奏の経験から、相手の演奏者の演奏の癖や特徴を学習し、この知識を利用してリハーサル前に相手に合わせた演奏プランの修正や構築を行なっていると考えられる。自動伴奏システムでも、過去の独奏者との演奏の履歴を利用することで、リハーサルのさらなる効率化や独奏者に特化した演奏プランの生成が可能であると考えられる。しかしながら、従来の伴奏システムはリハーサルや合奏で得られた演奏情報を、将来の合奏のために分析、保存を行なう枠組については全く研究されていない。この問題は、リハーサルの範疇を越えており、実現するためには合奏記録の保存だけでなく、演奏分析や演奏プランの自動生成を

行なう必要がある。

6.1.4 まとめ

本節では自動伴奏システムに関する研究について概観した。以上をまとめると、従来の自動伴奏システムは以下の能力が欠けている。

- 演奏プランの作成
- リハーサルから得た情報の応用能力
- 演奏者への適応

この問題を実現するためには以下の技術が必要である。

- 演奏者情報や合奏記録の蓄積
- 楽曲情報を利用した合奏記録の分析
- 演奏記録からの演奏プランの抽出と合成

次節では事例に基づく演奏表情生成手法を利用して、これらの問題を解決する方法を提案する。

6.2 事例に基づく演奏表情生成を用いた自動伴奏システム

従来の伴奏システムには演奏プランの作成やリハーサルにそれぞれ課題があるという点は前節で述べた通りである。従来の伴奏システムでは、演奏プランの作成とリハーサルを全く別個の手続きとして行っていた。しかし、これらの二つの手続きは、いずれも演奏プランに関係したものであり、お互いに深く依存しあっていると考えられる。従って、これらの二つの手続きを一つに統合した自動伴奏システムのモデルが望まれる。

本節では、自動伴奏システムに事例に基づく演奏表情生成手法を導入することにより、演奏プランの作成とリハーサルを統合した自動伴奏システムのモデル [47, 48] について述べる。

6.2.1 構成

本章で提案する自動伴奏システムのモデルの特徴は、事例に基づく演奏表情生成手法を利用して演奏プランを自動生成することにある。演奏プランは演奏表情の元となる情報であるため、その内容は演奏表情に近いものである。従って、事例に基づく演奏表情生成手法で演奏プランを生成することが可能である。事例に基づく演奏表情生成手法を導入した自動伴奏システムの基本構成を図 6.4 に示す。この自動伴奏システムは、以下の二つのモジュールで構成されている。

- 伴奏システム部

伴奏システム部は従来の自動伴奏システムと同等の機能を持った部分である。実際に合奏を行なう際には、この伴奏システム部が全ての処理を行なう。伴奏システム部の構成は従来の自動伴奏システムと同じである。

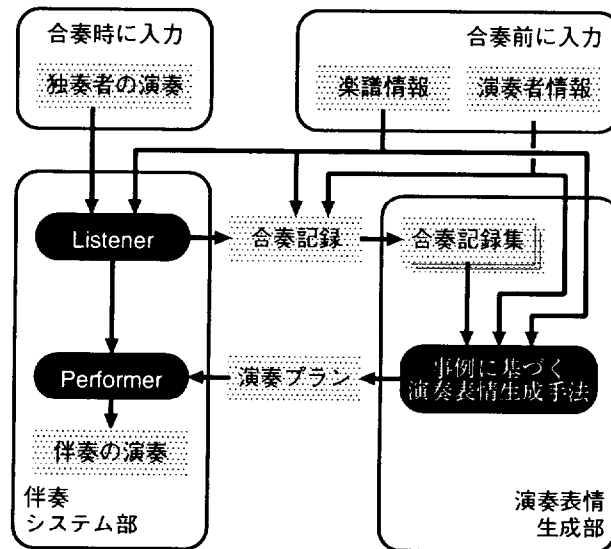


図 6.4: 事例に基づく演奏表情生成手法を導入した自動伴奏システム

- 演奏表情生成部

演奏表情生成部は事例に基づく演奏表情生成手法が用いられており、演奏前の演奏プランの生成やリハーサルの記録の処理を行なう。伴奏システム部で利用する演奏プランを過去の合奏の記録情報から生成する。

この二つのモジュールは演奏データと演奏プランを通して互いに接続されている。次に、個々のデータについて概説する。

- 入力

入力は、演奏者情報と楽譜情報からなる。演奏者情報は、独奏の演奏者が誰であるかを特定する情報である。この情報は後述の合奏記録集で演奏状況として利用する。楽譜情報は合奏曲の楽譜情報である。独奏部分と伴奏部分の楽譜情報をそれぞれ別個に入力する。また、楽曲構成情報も楽譜情報と共に入力する。

- 独奏者の演奏入力

演奏時の独奏者の演奏操作である。MIDI 信号を利用してリアルタイムに入力される。この演奏操作情報は Listener が独奏者と楽譜との照合に利用する。また、演奏操作は逐次記録し、後述の合奏記録とする。

- 合奏記録

合奏記録は、合奏曲の楽譜情報、合奏時の独奏者の演奏記録、独奏部分の演奏者情報からなる。楽譜情報と演奏者情報は、合奏時の入力と同じである。演奏記録は独奏者の演奏時の打鍵や離鍵などの演奏操作を記録した演奏データである。

- 合奏記録集

合奏記録集は自動伴奏システムが過去に行なった全ての合奏の記録を蓄積したデータベースである。

- 演奏プラン

演奏プランは、伴奏の演奏表情の元となる情報である。合奏前に演奏表情生成部が生成し、伴奏システム部に入力と共に与えられる。演奏プランは伴奏部分の一音ごとのテンポや音の強さなどの演奏表情値である。従って、内容は演奏表情情報と同様である。

6.2.2 システムの動作

この自動伴奏システムの基本動作を 図 6.5 に示す。このシステムの基本動作は、以下に示す一連の手続きの繰り返しである。

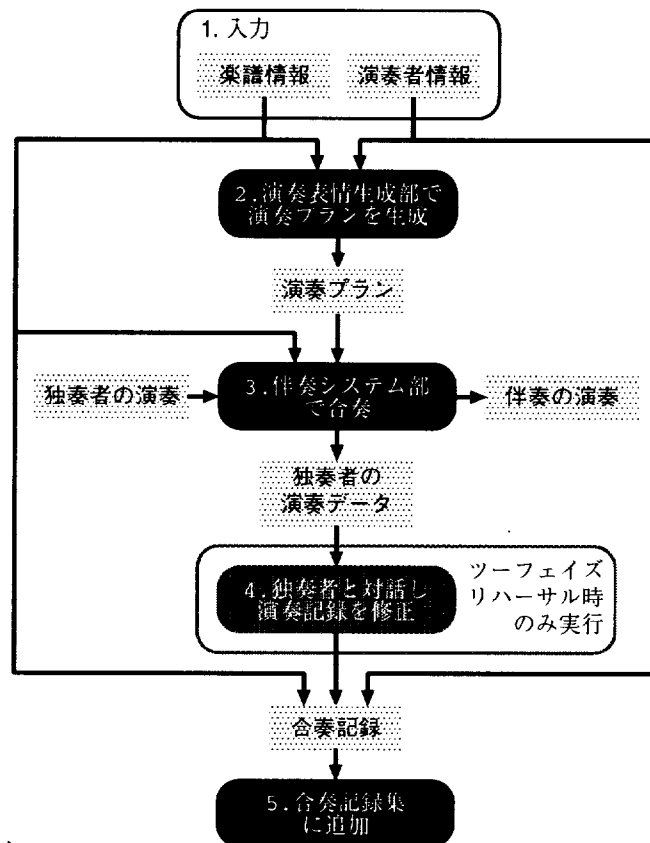


図 6.5: 事例に基づく演奏表情生成手法を導入した自動伴奏システムの動作

1. 入力

合奏を行なう前に演奏者情報と楽譜情報をシステムに与える。

2. 演奏プラン生成

入力された演奏対象曲の伴奏部分に対する演奏プランを演奏表情生成部で生成する。演奏プランの生成には事例に基づく演奏表情生成手法を利用する。自動伴奏システムで扱うデータと事例に基づく演奏表情生成手法に必要なデータとの関係を表 6.1 に示す。自動伴奏システムの演奏者情報は、事例に基づく演奏表情生成手法の演奏状況情報として利用する。事例に基づく演奏表情生成では、演奏データの作成は行なわず、演奏対象曲の一音ごとの演奏表情値が得られたところで停止する。この対象曲

の演奏表情値を伴奏システム部で演奏プランとして使用する。

事例に基づく演奏表情生成手法		伴奏システム
演奏データ集	楽譜情報	合奏記録の楽譜情報
	旋律断片境界	合奏記録の楽譜情報に付随
	演奏データ	合奏記録の独奏者の演奏データ
	演奏状況	合奏記録の演奏者情報
入力	演奏対象曲の楽譜情報	入力された楽譜情報の伴奏部分
	演奏対象曲の旋律断片境界	入力された楽譜情報に付随
	演奏状況設定	入力された演奏者情報
出力	一音ごとの演奏表情値	演奏プラン

表 6.1: 自動伴奏システムのデータと事例に基づく演奏表情生成手法のデータとの関係

3. 合奏

伴奏システム部を利用して合奏を行なう。その際に演奏表情生成部が生成した演奏プランを利用する。また、独奏者の演奏操作は逐次記録する。

4. コミュニケーション（リハーサルの場合）

この自動伴奏システムは堀内が提案したツーフエイズリハーサルを導入することもできる。ツーフエイズリハーサルを導入する場合は、この段階でコミュニケーションフェイズを行なう。ここでは、独奏者から与えられた要求に基づいて、直前の合奏の演奏記録を修正する。

5. 合奏記録を演奏データ集に追加

合奏の演奏記録、演奏者情報、楽譜情報を合奏記録として合奏記録集に追加する。コミュニケーションフェイズを行なった場合は、要求に基づいて修正した演奏記録を合奏記録として追加する。

通常の合奏の場合もリハーサルの場合も、この一連の手続きを行なう。リハーサルの前と後では、合奏記録集に含まれる演奏データが変化するため、演奏表情生成部が生成する演奏プランに違いが生じる。そこで、状況ごとのシステムの動作について演奏表情生成部の動作を中心に説明する。

独奏者が未知の場合 まず、独奏者 A と初めて合奏を行なう場合について説明する。初めての相手との合奏の場合、合奏記録集の中には演奏状況が独奏者 A となっている演奏データは含まれていない（図 6.6）。事例に基づく演奏表情生成手法では、4.4 節 で述べたように演奏状況の素性が一次独立であるとしているため、演奏プラン生成時には演奏者が異なる演奏データは全て演奏状況の類似性が 0 になる。従って、全ての合奏記録を対等に評価し、そこから標準的な演奏プランが生成される。

合奏経験がある相手の場合 過去に独奏者 A との合奏を行なったことがある場合、合奏記録集に独奏者 A との合奏の演奏データが含まれている（図 6.7）。事例に基づく演奏表情では演奏状況が同じ演奏データが重視されるため、独奏者 A との合奏の記録は他の独奏者との合奏データより重要度の高い演奏データとして扱われる。その結果、未知の独奏者との場合に比べて、独奏者 A の演奏の特徴を反映した演奏プランが生成される。

リハーサルを行なった場合 合奏のリハーサルを行なった場合、合奏記録集に入力された演奏曲と同じ楽曲の演奏データが含まれており、さらに、その合奏記録の演奏状況が独奏者と一致する（図 6.8）。5.4.1 項で述べた演奏表情生成システムのクローズドテストの結果からも明らかのように、対象曲と同じ楽曲の演

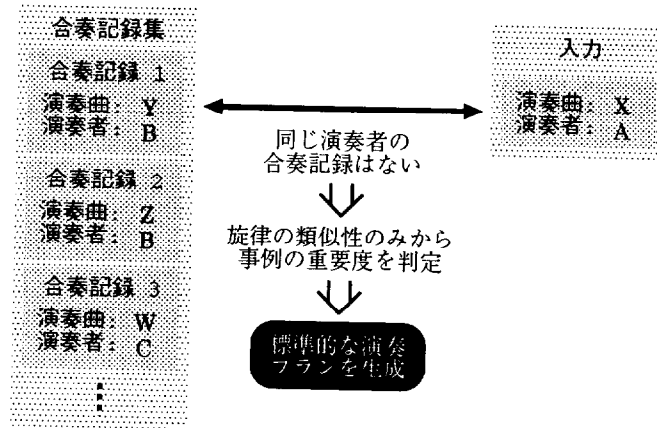


図 6.6: 初めての相手との合奏

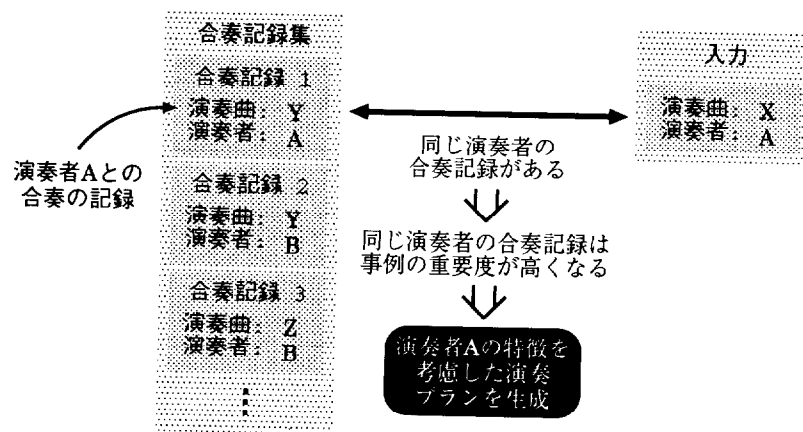


図 6.7: 演奏経験がある相手との合奏

奏データが演奏データ集に含まれており、演奏状況も入力と一致していた場合、演奏データ集の演奏データに極めて近い演奏表情が生成される。従って、リハーサル時の独奏者の演奏に近い演奏プランが生成される。

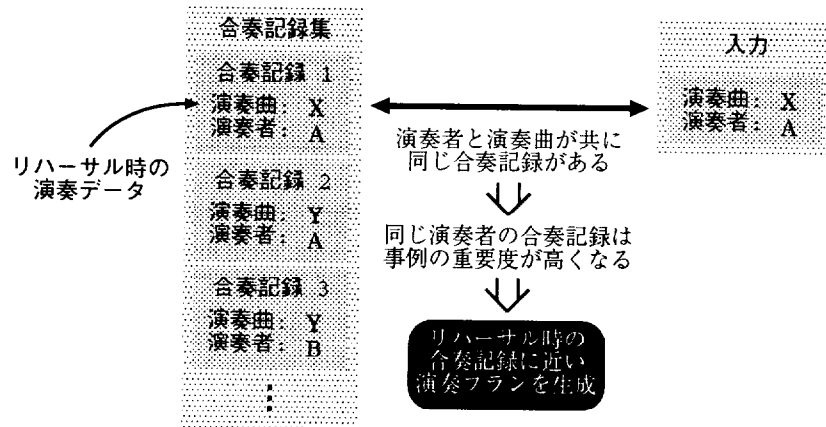


図 6.8: リハーサル後の合奏

このように、伴奏システムの演奏経験やその時々々の独奏者に応じて、適切な演奏プランを用意することが可能となる。

第7章 結論

7.1 まとめ

本論文では演奏状況に応じた多様な演奏表情の生成が可能な手法として、事例に基づく演奏表情の生成手法を提案した。この手法は、演奏表情に関する知識として、演奏表情の規則ではなく人間による演奏データを利用する。入力として対象曲と演奏状況設定を与えると、対象曲に類似した曲の演奏データを演奏データ集から検索する。そして検索された演奏データで人間が行っていた演奏を分析し、その演奏表情を対象曲に適用することで演奏表情の付与されたデータを生成する。この手法は、演奏表情の生成過程で演奏状況を考慮することが容易であるため、同じ楽曲に対して演奏状況の設定に応じた多様な演奏表情を生成することが可能であるという特徴をもつ。また、新たな演奏表情知識の追加の際に生じるコストが非常に小さい点や、演奏表情生成の際に認知的楽曲構造に関する知識が不必要であるといった点で、従来の規則に基づく手法に比べて優れている。

事例に基づく演奏表情の生成手法に関連した要素技術として、演奏データの有効利用に関する技術、演奏表情の表記・計算法、旋律の類似性の評価法、演奏状況の記述と類似性の評価の4つの技術を提案した。一般的に事例に基づく手法は検索によって得られる参考事例の数が不足していると有効に機能しない。事例に基づく演奏表情生成手法に於いて事例不足の問題の発生を軽減する技術として、楽曲を旋律断片に分割して検索する手法と、旋律断片の全体的な特徴のみに注目した検索手法について提案した。この技術を利用するためには、演奏表情の表記・計算法と旋律の類似性の評価に新たな手法を導入する必要がある。演奏表情の表記・計算法については、演奏表情の全体的な流れと局所的な揺らぎを分離することが可能な演奏表情の表現手法として演奏表情の大局比率を導入すると共に、この大局比率を用いた演奏表情の計算法などを示した。旋律の類似性の評価法については、旋律断片の全体的な特徴のみに注目し詳細な部分の違いの影響を受けにくいような類似性の評価方法を提案した。この旋律の類似性の評価法は参考事例の検索や重要度の計算の際に利用する。また、事例に基づく演奏表情生成手法の大きな特徴の一つである演奏状況の考慮に関連して、演奏状況の計算機上での表現法を提案すると共に、この表現法を用いた演奏状況同士の類似性の評価方法を示した。

事例に基づく演奏表情生成手法の応用として、自動演奏システムと伴奏システムについて概説した。自動演奏システムは、与えられた楽曲に対して、事例に基づく演奏表情生成手法を利用して、自然な演奏表情を付与し、その演奏データを生成するシステムである。このシステムの生成した演奏データに対して評価実験を行なったところ、事例に基づく演奏表情生成手法が自然な演奏表情の生成に有効であることや、演奏状況の設定に柔軟に対応し多様な演奏表情の生成を行なうことが可能であることなどが確認された。また、事例に基づく演奏表情の生成手法の伴奏システムへの導入についても検討を行なった。事例に基づく演奏表情生成手法は、伴奏システムにおいて課題となっているリハーサルや演奏者の癖の学習などに有効であると考えられる。

7.2 今後の課題

最後に今後検討すべき課題を以下にあげる。

- 旋律の類似性の評価手法の洗練
- 参考事例の検索の高速化
- 演奏状況についての考察
- 旋律の境界情報の与え方と演奏表情との関係の分析
- 和音を含む楽曲に対応した演奏システムの構築
- 事例に基づく演奏表情生成手法を導入した伴奏システムの評価実験

旋律の類似性の評価手法の洗練 事例に基づく演奏表情生成手法では、参考事例の検索や事例ごとの重要度の評価の際に、旋律断片の類似性の評価を行なっている。事例の重要度は最終的に生成されてくる演奏表情の決定に非常に大きな影響を与えるため、旋律断片の類似性の評価モデルが、生成される演奏表情の特徴や質に与える影響は非常に大きい。

旋律の類似性の評価手法については 4.3 節で比較的単純な手法を提案した。この手法を用いた演奏表情システム（第 5 章）の評価実験では良好な結果が得られていることから、この単純な手法でも演奏表情生成に対して有効に機能しているということができよう。しかしながら、この類似性方法自体を単独で評価したわけではない。この類似性評価手法の評価結果を分析することによって、様々な問題点や改善点が見つかる可能性も高い。

さらには、旋律の類似性の評価手法に学習機構を導入し、人間の演奏に近い演奏データを出力するような類似性評価モデルを学習させることにより、演奏表情生成システムに的した旋律の類似性評価モデルを獲得するなどの試み [44, 45] も有効であると思われる。

また、このようにして学習された旋律の類似性の評価モデルによる旋律の類似性の評価結果と、実際の人間による旋律の類似性の評価の間にはどのような関係が見られるかも非常に興味深い問題である。

参考事例の検索の高速化 第 5 章に示した演奏表情生成システムでは、参考事例の検索を行なう際に、対象曲を構成する個々の旋律断片について、演奏データ集の中にある全ての演奏データの個々の旋律断片との類似性を評価し、類似性の高い旋律を参考事例としている。この時の計算オーダは現状で $O((l \cdot \log l)^2 \cdot n)$ となる。 l は楽曲の平均長、 n は演奏データ集の曲数を表す。この計算量はかなり膨大であり、将来的に演奏データ集の総量や楽曲の長さが増化すると、検索に要する時間も相当なものになってしまうことが予想される。Kagurame Phase-I の生成実験 (5.3 節) では、5.4.4 項で述べたように、演奏データの生成に要する時間の 99% 以上を参考事例の検索で消費しており、実用的な演奏表情生成システムを構築するためには、この参考事例の検索に要する時間を如何に短縮するかが鍵となる。

少なくとも現状のように全ての旋律断片同士を比較するような手法では、計算オーダ的に苦しいことは明らかである。参考データ集の全ての旋律断片について、その特徴情報に従ってインデックスをふったりグルーピングするなどの前処理を行なうとともに、これらの情報を利用した計算オーダの少ない類似旋律検索アルゴリズムを構築する必要がある。これには旋律の検索技術に関する研究 [23] が有用であろう。

演奏状況についての考察 演奏状況については、4.4 節などで概説したが、まだまだ検討すべき点は多い。5.3 節の実験では演奏状況の素性として「ロマン派」と「古典派」の二つを利用した。これらの演奏状況は実験や評価を行なう上では非常に都合がよかったが、一般の楽曲に対応したシステムで利用するのに妥当であるかは判断が難しいところである。また、4.4 節などの例示で用いた演奏表情の素性にはあまり妥当な根拠がなく、ある意味で一般的なものを適当に選択したに過ぎない。演奏ごとの特徴を演奏状況で適切に表現するためには、演奏状況にどのような素性の集合を用いるべきかについて音楽心理学や音楽認知学 [34, 67] などの観点から検討する必要がある。

2.1.5 項では、演奏状況を「演奏表情に影響を及ぼす楽曲外の要因」と定義したが、実際には楽曲に関わる要因の中にも演奏状況として記述した方がよいものもある。例えば、単旋律の楽曲における和声などは、楽譜のみからでは一意に決定できないため演奏者の解釈に委ねられることがある。また、楽曲の作成された時期など楽曲に関する背景知識が演奏表情の決定に影響することも多い。このように、楽譜情報からは特定できないような楽曲に関する解釈や知識についても、演奏状況として考慮する必要がある。

4.4.3 項に示した演奏状況の類似性の評価手法では、素性は全て一次独立であるものと仮定している。しかしながら、素性の間にはお互いに密接な関係がある場合も多い。例えば、「明るい」と「暗い」という素性は、完全とは言い切れないにしても、対遇に近い関係にあることは明らかである。また、「はずむような」という素性には多分に「明るい」や「リズムカル」などと共通した特徴を持っているとも考えられる。このような演奏状況の素性同士の関係についても考慮する必要がある。

旋律の境界情報の与え方と演奏表情との関係 4.1.4 項で述べたように、本論文では計算コストの観点から、旋律断片の境界情報の一例として楽曲構成を利用することを提案した。また、第5章で紹介した演奏表情生成システムでも、この考えに基づいて旋律断片の境界情報を与えた。

しかしながら、類似旋律の検索が有効に行なわれるのであれば、必ずしも楽曲構成に基づいたものである必要はない。また、ここで利用している楽曲構成に基づく分割がどの程度有効な分割であるのかについても明らかではなく、他にも良好な結果が得られるような分割方法が存在する可能性が非常に高い。

そこで、4.1.2 項に示した理想的な分割手法や他の知識に基づく分割手法を用いて生成した演奏表情と比較することによって、楽曲構成に基づく分割がどの程度有効であるかを検証する必要がある。また、分割方法によって生成される演奏データのさらなる改善が期待できるのであれば、計算コストと演奏表情の双方の観点から良好な結果が得られる分割手法を追求する必要がある。

和音を含む楽曲に対応した演奏システムの構築 本論文で用いた実験システムは、システムの複雑化を避けるため単旋律の楽曲にしか対応できない。しかしながら、単旋律の楽曲では演奏表情の適切性の評価が難しいという意見があった。また、実用的なシステムとして利用するためには和音を含む一般の楽曲に対応することが必要不可欠である。

一般の楽曲に対応可能なシステムを構築するためには、和音を含む旋律断片同士の類似性を評価することが可能な、旋律の類似性の評価手法が必要となる。また、和音の各音の間の音の強さの比率などの情報を取得するため、旋律断片の分割を時間方向だけでなく、周波数方向（和音の分割など）にも拡張する必要があるかも知れない。さらには、グリッサンドなど和音の楽曲のみに生じる演奏表情への対応も必要である。

伴奏システムの評価実験 事例に基づく演奏表情を導入した自動伴奏システムの評価実験は十分には行っていない。リハーサルや演奏者適応の効用の評価は難しいが、心理実験などを通して事例に基づく演奏表情生成手法がこれらの課題にどのような効果をもたらしているかについて評価したい。

謝辞

本研究を行なうにあたり、終始適切な御指導、御鞭撻を頂きました指導教官の徳永健伸助教授に心より感謝します。田中穂積教授には研究の進め方などについて暖かく御指導していただきました。深く感謝致します。

堀内靖雄氏（現、千葉大学）には音楽情報処理の研究を行なうきっかけを与えていただきました。また研究会発表などの折にも適切な助言をいただきました。藤井敦氏（現、図書館情報大学）には論文のまとめ方や関連研究に関して有益な意見をいただきました。

奥井学氏、宮川和氏、宮本朋範氏、西田深志氏、木暮雅樹氏には、歴代の音楽グループの一員として研究に協力して頂きました。田中・徳永研究室のメンバーの皆様には研究討論会や輪講だけではなく、研究室での日常生活を通じて大変お世話になりました。特に、乾健太郎氏（現、九州工業大学）、白井清昭助手、今井宏樹氏、植木正裕氏、橋本泰一氏には有益な助言を頂きました。また、Timoth Baldwin 氏、Christoph Neumann 氏には英語の校正をして頂きました。田中研究室歴代の秘書の皆様には事務手続きなどで大変お世話になりました。

以上の皆様に心より感謝致します。

参考文献

- [1] 青野裕司, 片寄晴弘, 井口征士. 重回帰分析を用いた演奏ルールの抽出. 情報処理学会研究報告 95-MUS-11, Vol. 95, No. 74, pp. 1-6, 1995.
- [2] 青野裕司, 片寄晴弘, 井口征士. 重回帰分析を用いた演奏表現法の抽出. 情報処理学会論文誌, Vol. 38, No. 7, pp. 1473-1481, 1997.
- [3] 青野裕司. ジャムセッションシステム. bit 別冊, コンピュータと音楽の世界, pp. 283-305, 1998.
- [4] J. L. Arcos, R. L. de Mantaras, and X. Serra. SaxEx : a case-based reasoning system for generating expressive musical performances. In *Proceedings of the 1997 International Computer Music Conference*, pp. 329-336. International Computer Music Association, 1997.
- [5] International MIDI Association. *MIDI 1.0 Specification*, 1983.
- [6] D. Bainbridge, D. Cope, T. Crawford, and et al. *Melodic Similarity: Concepts, Procedures, and Applications*. Computing in Musicology 11. MIT Press, 1998.
- [7] J. J. Bloch and R. B. Dannenberg. Real-time computer accompaniment of keyboard performances. In *Proceedings of the 1985 International Computer Music Conference*, pp. 279-289. International Computer Music Association, 1985.
- [8] R. Bresin, G. De Poli, and A. Vidolin. Symbolic and sub-symbolic rules system for real time score performance. In *Proceedings of the 1992 International Computer Music Conference*, pp. 211-214. International Computer Music Association, 1992.
- [9] C. Chafe. Statistical pattern recognition for prediction of solo piano performance. In *Proceedings of the 1997 International Computer Music Conference*, pp. 145-148. International Computer Music Association, 1997.
- [10] R. B. Dannenberg. An on-line algorithm for real-time accompaniment. In *Proceedings of the 1984 International Computer Music Conference*, pp. 193-198. International Computer Music Association, 1984.
- [11] P. Desain and H. Honing. Tempo curves considered harmful. a critical review of the representation of timing in computer music. In *Proceedings of the 1991 International Computer Music Conference*, pp. 143-149. International Computer Music Association, 1991.
- [12] P. Desain and S. Vos. Auto correlation and the study of musical expression. In *Proceedings of the 1990 International Computer Music Conference*, pp. 357-360. International Computer Music Association, 1990.
- [13] A. Friberg. Generative rules for music performance: a formal description of a rule system. *Computer Music Journal*, Vol. 15, No. 2, pp. 56-71, 1991.

- [14] A. Friberg, L. Fryén, L. Bodin, and J. Sundberg. Performance rules for computer-controlled contemporary keyboard music. *Computer Music Journal*, Vol. 15, No. 2, pp. 49–55, 1991.
- [15] A. Friberg and J. Sundberg. A lisp environment for creating and applying rules for musical performance. In *Proceedings of the 1986 International Computer Music Conference*, pp. 1–3. International Computer Music Association, 1986.
- [16] L. Frydén and J. Sundberg. Performance rules for melodies. origin, functions, purposes. In *Proceedings of the 1984 International Computer Music Conference*, pp. 221–224. International Computer Music Association, 1984.
- [17] L. Frydén, J. Sundberg, and A. Askenfelt. From music score to sound. a rule system for musical performance of melodies. In *Proceedings of the 1982 International Computer Music Conference*, pp. 426–435. International Computer Music Association, 1982.
- [18] 後藤真考, 日高伊佐夫, 村岡洋一. すべてのプレーヤーが対等なジャズセッションシステム i. システムの全体後奏と分散環境での実装. 情報処理学会研究報告 96-MUS-14, Vol. 96, No. 19, pp. 21–28, 1996.
- [19] L. Grubb and R. B. Dannenberg. A stochastic method of tracking a vocal performer. In *Proceedings of the 197 International Computer Music Conference*, pp. 337–340. International Computer Music Association, 1997.
- [20] 波多野諠余 (編). 音楽の認知. 認知科学選書 12. 東京大学出版会, 1987.
- [21] 日高伊佐夫, 後藤真考, 村岡洋一. すべてのプレーヤーが対等なジャズセッションシステム ii. ベーシストとドラマーの実現. 情報処理学会研究報告 96-MUS-14, Vol. 96, No. 19, pp. 29–36, 1996.
- [22] 平賀瑠美. 演奏の表情付け. bit 別冊, コンピュータと音楽の世界, pp. 270–282, 1998.
- [23] Y. Hiraga. Content-based retrieval and indexing methods for music databases. In *Proceedings of the 1999 International Computer Music Conference*, pp. 299–302. International Computer Music Association, 1999.
- [24] H. Honing. Poco: An environment for analysing, modifying, and generating expression in music. In *Proceedings of the 1990 International Computer Music Conference*, pp. 364–367. International Computer Music Association, 1990.
- [25] Y. Horiuchi, H. Tanaka. A computer accompaniment system with independence. In *Proceedings of the 1993 International Computer Music Conference*, pp. 418–420. International Computer Music Association, 1993.
- [26] 堀内靖雄, 藤井敦, 田中穂積. 複数の人間と協調する演奏システム ~人間とコンピュータとのアンサンブル~. インタラクティブシステムとソフトウェアに関するワークショップ, レクチャーノート/ソフトウェア学7 「インタラクティブシステムとソフトウェア I」, pp. 279–286, 1993.
- [27] 堀内靖雄, 橋本周司. 伴奏システム. 情報処理, Vol. 35, No. 9, pp. 815–821, 1994.
- [28] 堀内靖雄, 奥井学, 鈴木泰山, 田中穂積. 伴奏システムのためのリハーサル. 情報処理学会研究報告 94-MUS-8, Vol. 94, No. 103, pp. 51–56, 1994.

- [29] 堀内靖雄, 奥井学, 鈴木泰山, 田中穂積. 伴奏システムのリハーサル ～人間とコンピュータとのアンサンブル(2)～. インタラクティブシステムとソフトウェアに関するワークショップ, レクチャーノート/ソフトウェア学9 「インタラクティブシステムとソフトウェア II」, pp. 115-123, 1994.
- [30] 堀内靖雄, 田中穂積. 自主性を持つ伴奏システム. 人工知能学会誌, Vol. 10, No. 1, pp. 72-79, 1995.
- [31] 堀内靖雄, 藤井敦, 田中穂積. 複数の人間と協調する演奏システム. コンピュータソフトウェア, Vol. 12, No. 5, pp. 63-71, 1995.
- [32] 堀内靖雄. 自動伴奏. bit 別冊, コンピュータと音楽の世界, pp. 252-269, 1998.
- [33] 五十嵐滋, 彌富あかね, 小川大典, 清野桂子. PSYCHE から: フレーズ表情の多角形, 音楽構造に基づく表情付け, 自動連弾システム. 情報処理学会研究報告 95-MUS-9, Vol. 95, No. 19, pp. 1-6, 1995.
- [34] 井口征士, 中村敏枝, 他. 感性情報処理. ヒューマンコミュニケーション工学シリーズ. オーム社, 1994.
- [35] 井上渉, 橋本周司, 大照完. 適応型歌声自動伴奏システム. 情報処理学会論文誌, Vol. 37, No. 1, pp. 31-38, 1996.
- [36] 石桁真礼生, 末吉保雄, 丸田昭三, 飯田隆, 金光威和雄, 飯沼信義. 楽典:理論と実習. 音楽之友社, 1965.
- [37] 彌富あかね, 平賀瑠美, 五十嵐滋. 演奏ルールを用いる演奏表情の構造的展開. 人工知能学会第11回全国大会講演論文集, pp. 280-282, 1997.
- [38] 片寄晴弘, 上符裕一, 松尾聡子, 井口征士. 音楽認知の計算モデル —二つのアプローチ—. 情報処理学会研究報告 99-MUS-25, Vol. 98, No. 47, pp. 51-58, 1998.
- [39] 小池宏幸, 平賀瑠美, 五十嵐滋, 水谷哲也, 塩雅之. アゴーギクルールのパラメータ値の自動決定システムの構築. 情報処理学会研究報告 99-MUS-31, Vol. 99, No. 68, pp. 49-54, 1999.
- [40] J. Kolodner. *Case-based Reasoning*. Morgan Kaufmann, 1993.
- [41] J. LaRue, 大宮真琴. スタイル・アナリシス 総合的様式分析—方法と範例. 音楽の友社, 1988.
- [42] F. Lerdahl and R. Jackendoff. *A Generative Theory of Tonal Music*. MIT Press, 1983.
- [43] 劉剣利, 平賀瑠美, 五十嵐滋. DAPHNE: フレージングと表情付けのための叙述的楽音分析 システム. 情報処理学会研究報告 99-MUS-25, Vol. 98, No. 47, pp. 43-50, 1998.
- [44] 宮川和. 音楽におけるフレーズ間の類似性の評価及び相互関係の構造化に関する研究. 卒業論文, 東京工業大学 工学部, 1996.
- [45] 宮本朋範. 旋律の類似度の自動評価に関する研究. 卒業論文, 東京工業大学 工学部, 1998.
- [46] E. Narmour. *Analysis and Cognition of Basic Melodic Structures*. The University of Chicago Press, 1990.
- [47] 西田深志, 鈴木泰山, 徳永健伸, 田中穂積. 事例に基づく演奏表情の生成に関する研究. 情報処理学会研究報告 99-MUS-29, Vol. 99, No. 16, pp. 55-62, 1999.
- [48] 西田深志. 事例ベースで生成された演奏表情を用いて自然な演奏を行なう伴奏システム. Master's thesis, 東京工業大学 大学院 情報理工学研究科, 1999.

- [49] 西原祐一, 小杉尚子, 紺谷精一, 山室雅司. ハミング検索用音楽データベースの分析 ～部分曲分割方法の改善に向けて～. 情報処理学会研究報告 99-MUS-32, Vol. 99, No. 84, pp. 7-12, 1999.
- [50] M. Nishijima and K. Watanabe. Interactive music composer based on neural networks. In *Proceedings of the 1992 International Computer Music Conference*, pp. 228-231. International Computer Music Association, 1992.
- [51] K. Noike, N. Takiguchi, T. Nose, Y. Kotani, and H. Nisimura. Automatic generation of expressive performance by using music structures. In *Proceedings of the 1992 International Computer Music Conference*, pp. 211-214. International Computer Music Association, 1992.
- [52] 野池賢二, 乾伸雄, 野瀬隆, 小谷善行. 演奏情報と楽譜情報の対からの演奏表情規則の獲得とその応用. 情報処理学会研究報告 99-MUS-26, Vol. 98, No. 74, pp. 109-114, 1998.
- [53] 小田安彦, 白川健一, 村上豊, 梶川嘉延, 野村康雄. 演奏者情報を加味したピアノの自動演奏システムの構築:～ニューラルネットワークによる局所部分に於ける演奏特徴の抽出. 情報処理学会研究報告 95-MUS-10, Vol. 95, pp. 7-12, 1995.
- [54] D. Oppenheim. *Standard MIDI Files 0.06*, 1988.
- [55] 関本陽子, 野瀬隆, 乾伸雄, 小谷善行, 西村怒彦. 演奏情報に関する楽曲の特徴抽出システムの作成. 情報処理学会研究報告 99-MUS-21, Vol. 97, No. 67, pp. 1-6, 1997.
- [56] 白川健一, 小田安彦, 熊谷俊行, 梶川嘉延, 野村康雄. 演奏者情報を加味したピアノの自動演奏システムの構築:～音符属性に着目した個人的解釈ルールの生成～. 情報処理学会研究報告 95-MUS-10, Vol. 95, pp. 1-6, 1995.
- [57] 園田智也, 後藤真孝, 村岡洋一. Www 上での歌声による曲検索システム. 情報処理学会研究報告 99-MUS-24, Vol. 98, No. 14, pp. 25-32, 1998.
- [58] T. Suzuki, T. Tokunaga, and H. Tanaka. A case based approach to the generation of musical expression. In *Proceedings of the sixteenth International Joint Conference on Artificial Intelligence*, pp. 642-648. International Joint Conference on Artificial Intelligence Inc., 1999.
- [59] 鈴木泰山. 楽譜情報を考慮した伴奏システムのためのピッチ抽出に関する研究. 卒業論文, 東京工業大学 工学部, 1995.
- [60] 鈴木泰山, 徳永健伸, 田中穂積. 事例ベースの演奏表情生成手法に関する研究. 情報処理学会研究報告 97-MUS-21, Vol. 97, No. 67, pp. 7-12, 1997.
- [61] 鈴木泰山, 宮本朋範, 西田深志, 徳永健伸, 田中穂積. Kagurame phase-i —事例ベースの演奏表情生成システム—. 情報処理学会研究報告 98-MUS-24, Vol. 98, No. 14, pp. 61-68, 1998.
- [62] 鈴木泰山, 西田深志, 徳永健伸, 田中穂積. 事例ベースの演奏表情生成システム — kagurame phase-i —. 情報処理学会第 57 回全国大会講演論文集第 2 分冊, pp. 13-14, 1998.
- [63] 鈴木泰山, 木暮雅樹, 徳永健伸, 田中穂積. 事例に基づく演奏表情の生成に関する研究. 情報処理学会研究報告 99-MUS-32, Vol. 99, No. 84, pp. 25-30, 1999.
- [64] 田口友康. 自動演奏と演奏モデル:演奏の芸術性を目指して. bit 別冊, コンピュータと音楽, pp. 42-53, 1987.

- [65] 竹内好宏, 保科洋. 演奏表情を知覚させる演奏変数の研究. 情報処理学会研究報告 99-MUS-15, Vol. 96, No. 53, pp. 7-12, 1996.
- [66] 竹内好宏. 楽曲の構造解析とその応用. bit 別冊, コンピュータと音楽の世界, pp. 224-240, 1998.
- [67] 寺西立年, 大串健吾, 宮崎謙一. 音楽の心理学 (下). 西村書店, 1987.
- [68] N. P. Todd. A model of expressing timing in tonal music. *Music Perception*, Vol. 3, pp. 33-58, 1985.
- [69] N. P. Todd. Computational modeling of musical expression. In *Proceedings of the 1990 International Computer Music Conference*, pp. 373-375. International Computer Music Association, 1990.
- [70] Y. Uwabu, H. Katayose, and S. Inokuchi. A structural analysis tool for expressive performance. In *Proceedings of the 1997 International Computer Music Conference*, pp. 121-124. International Computer Music Association, 1997.
- [71] 上符裕一, 青野裕司, 片寄晴弘, 井口征士. 演奏ルールの抽出について. 情報処理学会研究報告 96-MUS-15, Vol. 96, No. 53, pp. 79-84, 1996.
- [72] B. Vercoe. The syntehtic performer in the context of live performance. In *Proceedings of the 1984 International Computer Music Conference*, pp. 199-200. International Computer Music Association, 1984.
- [73] B. Vercoe and M. Puckette. Synthetic rehearsal: Training the synthetic performer. In *Proceedings of the 1985 International Computer Music Conference*, pp. 275-278. International Computer Music Association, 1985.
- [74] G. Widmer. The synergy of music theory and AI: Learning multi-level expressive interpretation. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, pp. 114-119. American Association for Artificial Intelligence, 1993.
- [75] G. Widmer. Understanding and learning musical expression. In *Proceedings of the 1993 International Computer Music Conference*, pp. 268-275. International Computer Music Association, 1993.
- [76] G. Widmer. Learning expression at multiple structural levels. In *Proceedings of the 1994 International Computer Music Conference*, pp. 95-101. International Computer Music Association, 1994.
- [77] G. Widmer. Modeling the rational basis of musical expression. *Computer Music Journal*, Vol. 19, No. 2, pp. 76-96, 1995.
- [78] G. Widmer. Qualitative perception modeling and intelligent musical learning. *Computer Music Journal*, Vol. 16, No. 2, pp. 51-68, 1995.
- [79] 矢向正人, 土屋景一, 荒木敏規. 旋律パターンの分類 -類似性と分析例-. 情報処理学会研究報告 99-MUS-16, Vol. 96, No. 75, pp. 27-32, 1996.
- [80] 山内敏彰, 野池賢二, 野瀬隆, 乾伸雄, 小谷善行, 西村怒彦. Ga を用いたピアノ演奏の局所テンポ解析. 情報処理学会研究報告 96-MUS-14, Vol. 96, No. 16, pp. 1-6, 1996.