



Title	音楽感性情報処理に関する研究
Author(s)	片寄, 晴弘
Citation	大阪大学, 1991, 博士論文
Version Type	VoR
URL	https://doi.org/10.11501/3054382
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

音楽感性情報処理に関する研究

片 寄 晴 弘

1991年1月

大阪大学基礎工学部

目 次

第1章 総論	-----	1
1-1 音楽情報処理	-----	1
1-1-1 音の生成技術	-----	2
1-1-2 ノーテーションシステム	-----	3
1-1-3 マンマシン・インターラクション	-----	4
1-1-4 楽譜の認識システム	-----	4
1-1-5 採譜システム	-----	5
1-1-6 作・編曲システム	-----	5
1-1-7 総合的音楽情報処理システム	-----	6
1-1-8 音楽データ	-----	7
1-2 感性情報処理	-----	9
1-2-1 関連領域との関係	-----	9
1-2-2 感性情報処理の対象	-----	12
1-3 本研究の意義と論文の構成	-----	15
第2章 音楽感性情報処理	-----	17
2-1 音楽感性情報処理	-----	17
2-1-1 音楽の認識と理解	-----	17
2-1-2 作曲プログラム	-----	21
2-2 音楽感性情報処理システム “KANSEI”	-----	23
2-2-1 低次音楽情報法処理	-----	25
2-2-2 高次音楽情報処理	-----	26
2-3 音楽データ	-----	27
2-3-1 scr 形式	-----	28
2-3-2 note 形式	-----	29
2-3-3 内部データ	-----	30
2-4 まとめ	-----	31
第3章 自動採譜	-----	33

3－1 自動採譜における諸問題	33
3－1－1 音響信号から音素シンボルへの変換過程	34
3－1－2 音素シンボルから楽譜への変換過程	39
3－2 複素スペクトルの内挿法と音階の単位系	39
3－2－1 複素スペクトルの内挿法	40
3－2－2 音階の単位系	41
3－3 多重音の採譜システム	43
3－3－1 システムの概要	43
3－3－2 各モジュールの説明	45
3－3－3 ビート・トラッキング	45
3－3－4 音名の同定	48
3－3－5 ノートデータから楽譜への変換	48
3－3－5－1 ノートデータの解析	49
3－3－5－2 楽譜シンボルの図形化、表示	51
3－3－6 処理手順および実験	52
3－3－6－1 ノートデータの抽出	53
3－3－6－2 楽譜生成結果	54
3－3－7 多重音の採譜システムのまとめ	57
3－4 複数音源の分離	58
3－4－1 アコースティック・アルゴリズム・コンパイラ	58
3－4－2 実験	60
3－4－2－1 楽器の推定	62
3－4－2－2 ピアノとクラリネット音の分離	63
3－4－3 検討	63
3－5 まとめ	66
 第4章 音楽分析と理解	67
4－1 音楽プリミティブ	67
4－1－1 リズムパターンの認識	69
4－1－2 調性の認識	69
4－1－3 メロディの抽出	70
4－1－4 コード進行の認識	71

4-1-5 音楽プリミティブと形容詞とのリンク	72
4-1-6 実験および検討	73
4-2 音楽聴取モデル	76
4-2-1 人間の音楽聴取の観察	76
4-2-2 音楽聴取モデル	77
4-3 感性状態の記述	78
4-3-1 伝播モデル	79
4-3-2 ファクタモデル	79
4-4 予期モデル	81
4-5 検討	84
4-6 まとめ	85
 第5章 音楽解釈	 86
5-1 自動演奏と音楽解釈	86
5-2 演奏解釈システムMIS	87
5-2-1 楽譜認識過程	90
5-2-2 演奏生成過程	91
5-2-3 演奏ルール学習過程	91
5-3 データ表現とルール	91
5-3-1 データ表現	91
5-3-2 データの内部表現	93
5-3-3 ルールの働き方	94
5-4 表現ルール	94
5-4-1 フレージング	95
5-4-2 モチーフの表現	98
5-4-3 モチーフ構造に働くルール	99
5-4-4 曲の種類が規定するモチーフ表現	99
5-4-5 表意記号の表現	100
5-4-6 感性情報の表現	101
5-5 楽曲の構造解析とグルーピングストラテジー	101
5-6 演奏分析	106
5-6-1 音楽表現の抽出	106

5-6-2	一段目の対応探索	108
5-6-3	二段目の対応探索	109
5-6-4	和音内の音の立ち上がり時刻、音の長さ・パワーの同定	111
5-7	表現ルールの学習	111
5-7-1	表意記号の演奏法の学習	112
5-7-2	モチーフの演奏法の学習	112
5-7-3	フレーズの演奏法の学習	115
5-8	グルーピングストラテジーの抽出	118
5-9	演奏ルールの抽出に関する検討	119
5-10	まとめ	120
第6章 作曲過程		121
6-1	作曲と創作について	121
6-2	作曲支援システムにおける感性情報の利用	122
6-2-1	作曲支援	123
6-2-2	編曲支援	124
6-2-3	メディアの変換	125
6-3	まとめ	125
第7章 結論		126
謝辞		129
参考文献		130
付録		136

音楽感性情報処理に関する研究

片寄 晴弘

第1章 総論

現代社会は科学技術の発達によりかつての歴史をひもといてもたぐいない成熟の期を迎えており、現代を生きる人々の指向はよりソフトで、そして精神的な方向に向かい一つある。この社会を支えて来た科学技術のうち最も大きなもののひとつに情報処理技術があるが、それに対して求められるものも社会動向とともに変遷しつつある。従来の情報処理の対象といえば数値情報であり客観的情報であったが、最近では、芸術あるいはよりよい機械とのインターフェースに対する技術が求められている。本章では、そのような状況のもとに近年情報処理の一分野として数えられるに至った音楽情報処理・感性情報処理について述べ、本論文の主題となる音楽感性情報処理への導入を行なう。

1-1 音楽情報処理

近年、コンピュータは工学的な分野にとどまらず、数値情報とは一見かけ離れた芸術分野においてもその利用が進んでいる。中でも音楽分野においては早くからコンピュータの利用が研究され、1957年にコンピュータによる自動作曲「イリアック組曲」が発表されている[Hiller 1959]。そして1969年には後の楽音合成用プログラミング言語に多大なる影響を与えるMusic Vが発表されている[Mathews 1969]。以降、音楽とコンピュータの接点における研究は自動演奏システム、作編曲システム、楽譜印刷システム、採譜・読譜システム、音楽解析支援システムなどコンピュータの進歩とともにその裾野を広げ、音楽サイドで定着したコンピュータ・ミュージックに対して、科学・技術サイドにおいて音楽情報処理という研究領域が確立するに至っている。現在では音楽と計算機をテーマとした国際学会も複数を数え、コンピュータ・ミュージックおよびそれを支える技術、

そして、音楽の科学的解析あるいは音楽認知の研究成果が発表されている。

この節ではさまざまな側面を見せる音楽情報処理のなかでも、情報処理という観点から音の生成技術、ノーテーションシステム、マンマシン・インターラクション、楽譜認識、自動採譜、作・編曲システム、総合的音楽情報処理システム、音楽データについて述べる。

1-1-1 音の生成技術（シンセサイザとシーケンサ）

現在までに最もコンピュータ・ミュージックの発展に貢献してきた技術は、やはり、シンセサイザとそれをコントロールするためのシーケンサであろう。

シンセサイザとはもともと合成を行なう機械（もの）という意味の言葉であるが、電子楽器としてのシンセサイザの方がもはや一般的となっている。

時代を振り返るとシンセサイザという楽器が定着したのは70年代である。70年代はアナログシンセサイザーの時代であり、当時プロに使われていたシンセサイザーはアナログコンピュータとほぼ同種のものであった。70年代は楽音の物理的モデルの研究が進んだ年代でもある[Nakamura 1981]。78年には、より操作性のよい音楽合成のプログラムという目的を持ったMUSIC11が発表されている[Vercoe 1979]。MUSICVでは同時に発音できる数と使用できるメモリ量が限られていたが（32音で、各音35ワードのメモリ量），MUSIC11は各音に可変長型のメモリを割り当て、メモリが不足するときはガーベジコレクションを実行することによって柔軟性を実現している。80年代以降はデジタル技術の発達、メモリの低価格化に支えられ、デジタル音源の時代に入った。80年代になってFM方式[Cowining 1973]を用いたシンセサイザがその音質の良さからひろく使われたが、最近ではサンプラーに代表されるように音響の波形データを直接記録・再生するものが主流となりつつある。ハードウエアでは音の入出力機能をもったホビー指向のパソコン・コンピュータ、ソフトウエアでは時間領域、周波数領域の双方で音響信号を加工するツールも開発されており、今後デジタル音源の利用はますます進むものと思われる。

一方、シーケンサについては、初期のものは特定楽器の自動演奏を目的にハードウェアとして市販されていたものであった。その後、MIDI（異なったメーカーのシンセサイザを同一のプロトコルで制御するために制定された音楽演奏用インターフェース）の標準化が行なわれ、シーケンサ、コンピュータ・ミュージックは大きな転換期を迎えた

[International MIDI Association 1983]. ハードウェアとしてのシーケンサのほかにパーソナル・コンピュータを用いた多機能のソフトウエア・シーケンサが数多く開発され、プロの音楽製作の方法が一変した。それとともにコンピュータ・ミュージックのアマチュアレベルにおける普及が進んだ。楽器店等で一台のパーソナル・コンピュータでドラム、ギター、ベース、ストリングスなどの複数音源を制御し、自動演奏が行なわれているのもその例である。

最近では、オブジェクト指向に基づいたもの、リアルタイム性を主眼としたものなどそれぞれの音楽制作ニーズに応じたシーケンサが開発されている。

現在、MIDIの問題点としては、シリアル転送を基本としてデータ転送を行なっているために、大量の音楽データを送るときにどうしても演奏のずれ（時間遅れ）が発生してしまうこということがある。この問題に対してはベンチャー・ビジネスを中心に専用のI/Oが発売されている。その他にも電子的音源ではアコースティック楽器音のディテールを完全にシミュレートすることはできていないという問題もあるが、ここに述べてきたようなエレクトロニクス技術が音楽界に与えた影響は計りしれないものがある。

1-1-2 ノーテーションシステム

コンピュータの利用は音の生成だけではなくパブリッシングの分野（楽譜生成）でも進んでいる。一般に楽譜生成と音楽データベースとしての目的を持つシステムをノーテーションシステムという。楽譜生成に関する知識は音楽という分野にしては比較的明示的であり、ベンチャー・ビジネスを中心に非常に積極的に商品化が進められた[Byrd 1977]。この背景にはデザイン支援システムとして進められてきたグラフィックス機能、ヒューマンインターフェースの技術がある。これについては、1-1-7節、音楽データで言及する。

通常の楽譜生成に関しては、人間のエキスパートに引けをとらないシステムも紹介されている[Oka 1987]。一方で、譜割り・位置決めなど感性に訴える要素に関しては、微妙な知識の明文化が難しく人間に及ばない点もある。この点を補うものとしては人間の介在を前提にエディティング機能を充実されたものが開発されている。

1-1-3 マンマシン・インターラクション

ここ数年話題となっているのがマンマシン・インターラクションである。音楽分野におけるマンマシン・インターラクションはもともと音楽のライブ性・リアルタイム性をコンピュータで実現したいという動機から始まったものであり、人間の動きを各種センサーで検出し、その信号を演奏のトリガー信号とする新世代の楽器が開発されている[Chadabe 1984]。また、イベント（MIDIのデータ）の加工機能を組み込んだプログラマブルなMIDIコントローラは、今までなかった表現を産み出すツールとして視覚・聴覚的効果も高く、コンピュータ・ミュージックのライブ演奏において注目を集めている[Macover 1989]。ビジョンと組み合わせた研究としては、人間の指揮者の指揮棒の動きを画像処理で検出し、その動きで演奏を制御するシステムも開発されている[Morita 1989]。

自動伴奏システムは、伴奏者であるコンピュータをソリストである人間に追従させることを目標とした非常に興味深いシステムである[Dannenberg 1984, Vercoe 1984, Danneberg 1988]。コンピュータは、予め、ソロのパートと伴奏パートを知っていて、ソリストがどの部分を弾いているかの監視を続けながら、リアルタイムで伴奏情報をシンセサイザに伝える。したがって、ソリストがつけたテンポの変化、休止時間などにほとんど違和感なく追従することができる。この研究の最も興味ある点はソリストの演奏のトラッキングである。人間の演奏には楽譜からだけでは予想しきれない解釈（例えば、トリルの回数）やミスなどがふくまれているため、それらに対処するための手法が各地で研究されている。

1-1-4 楽譜の認識システム

楽譜の認識システムは、音楽データベース作成、あるいは、演奏データの入力という目的からその実用化が熱望されてきた。しかし、シンボルの接続や途切れあるいは冗長な表現法が意味解析に大きく影響するなどパターン認識上の困難があり、音楽情報処理の中では比較的後発の領域である。初期の楽譜認識の研究においては、対象を单一パートのメロディなどの簡単なものに限定し、プロダクション・プロフィール等の基本的なパターン認識手法を用いた楽譜認識が行なわれていた[Nakamura 1979]。このような楽譜認識システムの中でもリアルタイムでの処理を目指したものとしては、筑波博で公開さ

れたWABOTの視覚系を挙げることができる[Ohteru 1985]。WABOTは楽譜をカメラで読んで演奏を行なうロボットであり、記憶に残している人もいることだろう。

最近では、より複雑な（一般的な）楽譜を対象とすることを目的とした楽譜認識システムが研究されている。音楽的知識の利用したり[Andronico 1982, Kato 1990]、並列処理[Vercoe 1989]を用いることにより、従来の認識システムでは考えられなかった処理能力を持つシステムの研究・開発が進められている。

1-1-5 採譜システム

楽譜認識システムが視覚からのデータの収集を目的にしているのに対し、聴覚を通して音楽データを収集する目的をもっているのが自動採譜システムである。採譜における最も基本的な処理は信号処理である。そのため、初期の採譜システムの研究においては信号処理の手法の開発が中心的な課題であった。信号処理の限界もあって、初期の採譜システムの対象は簡単な単一のメロディであった。現在では、単音から複音へ、単一楽器から、複数楽器へというように対象に対する制約を軽減する方向で研究が進んでいる。それにしたがって、採譜システムの実現においては、信号処理だけではなく、音楽認識のモデル化が非常に重要であることが明らかにされてきた[Katayose 1990a]。モデルとしては、信号の知覚・認識のレベルのものと、より高次の音楽処理のものに大別することができる。前者の例としては、音源分離、拍の割当（リズムの認識）などを挙げることができる。後者では、コードの認識など音楽がどのように処理されるのかが中心的な課題となる。なお、音楽の認識のモデリングについては、本論文の中心課題と深くかかわっているため、第4章で取り上げる。

1-1-6 作・編曲システム

音楽分野で最も早くコンピュータが利用されたのは、冒頭でも述べたように自動作曲である。以来、自動作曲は各地で試みられたが、多くはコンピュータ上に音楽の秩序を与えるフィルタを構成し、初期値としてモチーフ、入力としてランダム信号等を与え、出力として楽曲を出力するというものであった。これらのアプローチでつくられた作品は、結果的には音楽性・オリジナリティという点で人間に及ぶものではない。創造の難

しさは、良い音楽とは何か、音楽が似ているとはどういうことかなどを（自立的に）コンピュータに理解させる難しさの裏返しである。また創作は人間が行うべきだという認識もあって、自動作曲というより作編曲支援環境としての研究が進み、一部には商品化に至っているものもある。

初期の作曲支援システムとは一種の楽譜エディタと自動ピアノを組み合わせたようなものであった。現在ではシーケンサとシンセサイザとの組み合わせで実現されているが、それだけでは作曲という知的作業を支援するには能力不足である。人間の作曲助手であれば、例えば第1転回の長7音を削除することを要請されると、それが当てはまるすべての箇所を修正したうえで、楽譜を書いたり、あるいは演奏したりすることができる。このような作業の前提となるものはシンボルの音楽的解釈である。この点に着目して実現された作曲支援システムとしてはRoads, Cointeなどの研究がある[Roads 1985b, Cointe 1983]。

1-1-7 総合的音楽情報処理システム

上記のような音楽情報処理技術がある程度確立されてくるにしたがって、それぞれの機能を統合化した総合的音楽情報処理システムの開発が進められるようになった。総合音楽情報処理システムは、（ユーザの）音楽処理の環境の提供を第一の目的に始まつたものであるが、各種のデータ変換機能が人間特有の知的処理に相当し、一種の人工知能システムと位置づけられるものもある。

伝統的に確立されてきた音楽製作においては、作曲者が基本的な楽譜を提供し、演奏者が解釈を行ない、また、アレンジャーがその曲のアレンジを行なうというようにいわゆる分業が行なわれていたが、現代の音楽製作においてはその領域わけがなくなりつつある。

総合的な音楽情報処理システム構築の代表的な立場の一つは、このような音楽製作スタイルにおける支援システムとして、システムの統合を進めていくことである。これに分類されるものとしてはビジネスとして開発が進んだものが多く、典型的な例としてはフェアライト社の音源、シーケンサ、楽譜生成を統合化した音楽システムや、近年盛んに開発されているアップル社のマッキントッシュの一連の音楽ソフトウェアツール群を挙げることができる。また、音楽活動において要求されるリアルタイム制御、グラフィック機能、オブジェクト指向言語などをツールボックスとして提供するワークス

テーションも開発されている[Rodet 1988]。国内において前者の思想を強くもったものとしては、東京大学のICOToneのプロジェクトを挙げることができる[Hirata 1987]。

一方、明確に前者と一線を画すものではないが、高次の音楽活動の自動化という点で人工知能システムとしての色彩が総合的音楽情報処理システムがある。これについては、国内において活発に研究が進んでいる。代表的なものとしては、本研究の主題となる感性（KANSEI）音楽情報処理システム、早稲田大学のグループが進めているシステム、筑波大学のPSCYCHEなどがある。感性（KANSEI）音楽情報処理システムでは、音楽の「聞く」「読む」「演奏する」「書く」機能のほかに、高次の音楽理解として「鑑賞する」「作曲する」「解釈する」などのモデル化を実現している[Katayose 1989]。早稲田大学のグループは、WABOTの視覚の開発以来、前述の指揮の認識処理などマシン・ビジョンを積極的に取り入れた総合的音楽システムの開発を行なっている[Ohteru 1989]。PSCYCHEは自動演奏という目的を中心に構築された音楽システムであり、演奏データのエディティング、音楽解釈の機能が中心となっている[Tomitsuka 1988]。このほか、ハミングを入力して、好みのジャンルに応じたアレンジを行なうコンピュータシステムが市販にいたっている[Fujii 1989]。

1-1-8 音楽データ

前節までにも述べてきたように、音楽情報処理は様々な側面を持っている。そして音楽データもどの側面から捉えるかによっていくつかの種類に分けて考えることができる。代表的な音楽データの分類を図1-1に示す[Oka 1990]。

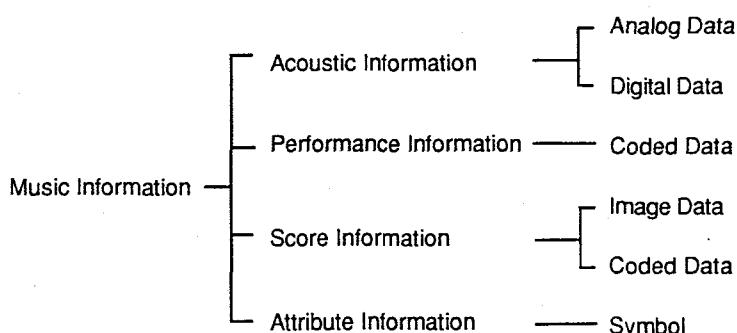


図1-1 音楽データの分類

これからもわかるように、音楽情報処理で対象とされるデータの範囲は非常に大きい。音楽情報処理における多くの分野では、それぞれの目的に応じて、データ記述に対する規則を制定してきた。この中で事実上の標準となっているデータ記述法もある。

音楽情報のうち演奏情報については、MIDIが制御信号レベルにおける規則を制定している。また、現存のコンピュータミュージックはMIDIを基本としているものが多く、各種のシーケンサソフトウェアで使われる演奏情報はMIDI信号の形で通信することができる。しかしながら、各ソウトゥエアで使われる高級演奏言語は基本的に互換性が補償されているものではない。次に、楽譜情報であるが、これについてはポストスクリプトのように電子式文書の記述法に従った音楽表記法がメーカーを中心に開発が進められている。

このような中で、音楽データの標準化を行なう検討がANSIの中で進められつつある。これは、音楽データの形式を統一することで、データの流通を活性化することを目的に始められたものである。中間報告として、ANSIのグループの出した規格案に対してのディスカッションが1990年のICMCでもたれたが、現段階では様々な意見が錯綜しており、今後の動向が注目されるところである。

一般に、音楽情報の標準化を図る立場としては、最初から、全ての立場からのニーズを満足するようなものを作る立場と、それぞれのニーズに応じて出来上がったデータ形式を組み入れてから標準化を行なうという立場がある。ANSIのグループはどちらかと言えば前者の立場に立っており、それが、標準化における大きな困難となっているようである。現状では、それぞれの固有領域における音楽データ、例えば作曲という固有領域一つを取ってもニーズは多様である。このような状況では、それぞれの領域におけるデータの標準化を行ない、その反応を見ることが必要であろう。それにもまして、変換可能なデータ間の変換可能技術（例えば音響データから楽譜データなど）を拡充することが重要である。

音楽情報の表記を考えるときにはその構造性を考慮していく必要がある。音楽とは時間的構造性の秩序の上に成り立つ芸術であるため、音楽表記においてもそれを反映することが望ましい。このことが音楽学的研究にも役立ち、そしてヒューマン・インターフェースの向上にも役立つはずである。このような観点で提案された音楽表記のうち、楽譜の構成を意識したものとしては、早稲田大学のSMX、ピアノの自動演奏における構造的な音楽表現を目指したものとしてはMUSEなどがある[Taguti 1989]。

1-2 感性情報処理

現在の高度情報化社会を支えている情報科学の枠組みは、論理的・客観的知識などのシンボリック情報の利用に基づいている。人工知能はその操作を、パターン認識は外界情報からその抽出を対象とし、それぞれ情報化社会を支える基盤技術として幅広い分野に大きく貢献してきた。論理的・客観的知識情報の社会的有効性は確かに誰もが認めるところであるが、我々人間が日常の社会生活において視覚や聴覚から直接的に得る情報と比べるとかなり縮約したものである。「美しい」「楽しい」「もう少しスマートに」というような形容詞的情報やそのような言葉で代表されるパターン情報は、冗長である、あるいは、人間の主観的な感覚に起因しコンピュータにはなじまない、という理由から従来の情報処理では除外されてきた。しかし、このような情報は人間同志円滑なコミュニケーションを行ううえで非常に重要な役割を果たしており、人間と機械との協調関係を作り出していくためには是非とも取り上げて行かなければならぬものである。また、物質的なものからより精神的なものに対して嗜好が強まりつつある現代社会においては、芸術、デザインといった創作処理におけるコンピュータ支援のニーズも大きく、その最終的な形としては、コンピュータが人間の感性を理解していることが望まれる。このような状況の下に感性を情報処理的な側面から議論し、機械と人間のよりよいインターフェースを実現することを目的とした感性情報処理という領域が確立しつつある。ここでは、感性情報処理の位置づけ、対象について述べる。

1-2-1 関連領域との関係

感性情報処理の関連領域として近年活発に取り組まれているものには、人工知能、心理学、認知科学などがある。この節では、これら関連領域と感性情報処理のかかわりについて考察を行なう。表1-1に感性情報処理、心理学、人工知能、認知科学のそれぞれの目的、対象について示す。人工知能にしても、認知科学にしてもその定義についてはさまざまな考え方があるだろうし、また各領域とも意味の基準がいつの時代でも同じであるとは限らない。この表は大勢的な見方の一つであることを断っておく。

表1-1 感性情報処理と関連領域の関係

分野	感性情報処理	心理学	領域	
			人工知能(広義)	
			人工知能(狭義)	認知科学
分野	工学	科学	工学	科学
目的	感性の機械化 情報の感性的側面を抽出し、記述し、生成する技術やシステムを提供する	感性の解明 感性(知能も含む)を、個々の事象の観察や実験に基づいて、解明する	知能の機械化 知覚、学習、判断、思考、記憶などの過程を人工的に実現し、人間生活に役立つ技術やシステムを提供する	知能の解明 知覚、学習、判断、思考、記憶などの過程を総括的にとらえ、その根本となる機構を解明する
主要な対象	感性情報	感性	知識	知能
計算機	主体	データの整理に利用	主体	モデルの検証に利用

「知」や「心」に対する研究アプローチは1940年代以降発展してきた情報科学とともに大きな変遷を迎え、情報機能との等価性という観点から議論されるようになってきた。そして、現在では「知能の解明」や「知能の機械化」を対象とした人工知能が社会的に見ても大きな研究対象として認知されるに至っている。特に、後者についてはニーズにも支えられた領域でもあったため、人工知能研究の中心的な対象として多くの研究者によって取り組まれてきた。その結果、後者を人工知能とする立場をとっている人も多い。ここでは後者に当たる領域を人工知能(狭義)としている。人工知能(狭義)の目的は「機械の知」を実現することである。したがって、人間の「知」を解明すること以上に知的機械を作ることが重要視され、科学というより工学的色彩の強い領域となっている。人工知能(狭義)における科学的目的は扱いうる世界と問題解決における入力と出力とを定式化するという点に向けられる。これに対し、認知科学は「人間の知」を理解することを目的とした領域であり、人工知能(狭義)以上に対象に関する観察と洞察に基づいたホワイトボックス的なモデリングが必要となる。この作業は、明らかに文科的な取り組みが必要である。しかしながら、最初に述べたようにコンピュータは情報機能のとしての等価性を測定を行なうためのツールとなりえたことによって、より数理モデル的な視点から研究が進められていくようになった。次に、人工知能(狭義)と認知科学の関係であるが、お互いに刺激を与えあい発展していると同時に関連した領

域を対象を異なった目的から取り組んでいるため、ある種の緊張感が見られることも事実である。認知科学側からは「人工知能（狭義）の研究は最も大切な考察をおざりにして、都合のよい世界でシステムを作っているだけである」というような意見が上がりたり、その逆に「認知科学がモデル化した対象があまりにも抽象的で、解決した問題などほとんどない」という声が人工知能（狭義）側から上がることもある。

感性情報処理は上記のような観点から「機械の知」の実現を目的とした人工知能（狭義）と非常に似た側面をもつ領域である。感性情報処理の目的を一言で言うと「機械の感性」を実現することである。「感性」とは「知能」以上に人間の本質に根ざした概念であり、システムを実現していくうえで人工知能（狭義）以上に人間の観察が必要であることは誰もが感じるところである。認知科学分野からは「感性を安易に考え過ぎていいのではないか」との危惧感も訴えられることがある。しかし、この危惧感の原因をよく考えると、感性情報処理に対する過度の期待、あるいは目的に対する誤解から生まれているものである。感性情報処理の主題は、より自然な形のインターフェースを実現するために、もう少し人間に歩み寄った機械を作るということであり、直接的には人間の感性を解き明かすことは目的としていない。そのため、具体的に感性情報を担う入出力を決定し、コンピューテーション・モデルを実現してゆくことが中心作業として行なわれる。その結果、ホワイトボックス的に見ると人間の心のモデルとは異なったモデルが構築されることもあるが、これは、人間の心とコンピュータのハードウェアが異なるからだ。当然のことであろう。

最近の認知科学の成果として、興味深いものに仮想的人格を計算機上に実動化を目指した“ねねプロジェクト”がある[Toda 1985]。“ねねプロジェクト”は、刺激と反応を会話に位置づけ、人格を感情をつかさどる複数のエージェントの総合的な合議結果であるというモデルをシステム的実現から考証しようとするものである。これは心理学、認知科学の興味から進められたプロジェクトであるが、システムの実現という点からは感性情報処理システムの具体例としてあげることができる。

感性情報を扱っていくときには従来の人工知能システムで用いらてきたシンボルや知識といった表現法以外の情報記述法、処理を確立してゆくことが必要である。これは感性情報処理という領域が将来的に取り組んでいく課題であり、そのアプローチについては本論文でも扱う。しかし、一旦、機械内部表現された情報については人工知能システムにおけるシンボルや知識の扱いと同一視できるものが少なくない。このことは、今までに提案してきた人工知能システムの多くの蓄積が利用できるという意味において非常に重要なことである。感性情報処理システムにとっては、人工知能システム以上に、

外界との情報の入出力すなわちインターフェースに対する取り組みが重要である。通常の人工知能システムの多くはキーボードを通じた情報を対象としてきたが、感性情報となるとキーボードを通じてやり取りで伝達できる情報より、五感を通じて直接的にやり取りされる情報の量が圧倒的に多いと考えられる。したがって、物理量と感覺量の関係測定やその手法の開発[Musha 1989, Uchio 1989]やもう少しソフトウエア的なところではパターンの生成や認識といった技術が非常に重要な研究対象となる。具体的には、話し方、顔つきから感性情報を認識したり音楽や絵画における表情を認識する機能、話し方や顔つきに表情を付けたりするなどの技術が非常に重要な研究対象となるが、これらのうち顔の表情の記述・認識・生成に関しては、知的情報通信という題材ですでに積極的に取り組まれている[Mase 1990, Choi 1991, Harashima 1989].

以上、感性情報処理と関連領域との関係を相違点を中心に述べてきたが、もちろん各領域は相異点以上に共通点をもった関連領域であることは忘れてはならない。システム作りと認知科学の興味の双方の興味から取り組まれてきた非単調論理の研究はその一例であろう。同様のことは、感性情報処理と関連領域についても予想されるところである。外界の情報から心的な感性情報に至るまでの過程においては、客観的な事象の認識・理解が必要なわけで、そのような点からは、感性情報処理を応用人工知能の研究領域として位置づけることもできる。また、(実験) 心理学における成果および方法論は感性情報処理システムの構築において不可欠なものであり、その結果は心理学に対して、新たなデータを提供することも期待される。

1-2-2 感性情報処理の対象

前節では、関連領域との関係という観点から、感性情報処理について述べてきた。ここでは、感性情報処理そのものの具体的な対象について述べる。

感性という言葉は、本来、印象を受け入れる能力としての意味を持つとされるが、拡大解釈された結果、「言葉で説明することは難しいが何か心に訴えるものがある」時に用いられるある意味で非常に便利な概念となっている。デザインなどの領域においては心理学実験を用いて個別に心理量の扱いが試みられてきたが、その一方で、言葉の持つ曖昧さがゆえに工学的に取り組まることはほとんどなかった。感性情報処理を議論していくためには、感性に対して、言葉の定義と扱うべき対象を明らかにする必要がある。ここでは、次のように感性を定義する。

(1) パターン感性情報

物理量でありながら、同定が難しいために“感性”の一言で置き換えられてきたパターン情報そのもの、例えば、ものの質感、音色、音楽の表現など

(2) シンボル感性情報

パターン感性情報は、物理量であると同時に主観的意味を持っており、中には特定の形容詞と密接に結びついたものもある。ここでは、パターン感性情報と密接に結びついた形容詞をシンボルとしてとらえ、シンボル感性情報とよぶことにする。主観量をシンボル的に扱うことには限界・問題点もあるが、人間とのインターフェースを考える時には、形容詞をキーとして利用することは有効である。ここでいう感性情報は、この観点から用いるものである。例えば、「暗い」=明度の低さ=音楽におけるマイナーなど。

(3) パラメータ感性情報

シンボル感性情報はパターン感性情報を表象する形容詞的シンボルであるが、すべてのパターン感性情報が特定のシンボル感性情報で表象されるとは限らない。このよう場合、パターン感性情報は、定義された概念（形容詞）空間において、一座標として表現することを考える。このとき、パターン感性情報や個々の形容詞は、複数のパラメータ（感性パラメータ）をもって記述される。

(4) イメージ感性

人間の心のシステムに深く依存し、イメージとして働くような主観、あるいは、直感。

イメージ感性に対して情報とつけなかったのは、現在、情報として扱う段階ではないと考えるからである。これは認知科学がこれから時間をかけて解き明かしてゆくべき問題そのものである。今後、心に関する研究が進んで完全な心理モデルが出来上がったとしても、実動のコンピュテーションシステムとするという制約があるため、現在のコンピュータのハードウエアに即した処理モデルを構築して行くことは不可欠である。現段階では、1と2そして3に分類される感性情報の記述と処理について研究して行くことが感性情報処理の切り口であると考えている。

パラメータ感性情報において感性情報（概念）がある空間内の座標によって表せることができるという立場は、心理学でいうとOsgood学派の考え方にも属するものである[Iwashita 1979]。感性情報（概念）が、ある空間内の座標で表現できる条件としては、その概念単位間の距離が測定できる必要がある。これに対しては、心理学においても論争が続けれられており、その結論は未だに見ないが、一方で、商品イメージの解析やデザインの方向付けなど実践面において有効性は確認されている。

1. パターン感性情報の抽出
2. パターン感性情報のパラメータ化, および, シンボル的感性情報との対応決定
3. パラメータ感性情報の操作・処理
4. パターン感性情報の生成
5. 各感性情報のデータベース化とその利用
6. イメージ感性量を扱うためのモデル化

感性情報処理の具体的な応用システムは、大きく次の二つに分類することができる。まずは、商品開発やデザインに対する総合的支援システムが上げられる。例えば、シンセサイザの音源を生成するときにもう少し、丸くとかいった作業が要求されるが、従来、このようなデザイン作業は、各企業あるいは研究機関のノウハウとして、職人的に蓄積されてきたものである。そのノウハウ自体の獲得が非常に煩雑であることを考えると、一旦データベース化されたノウハウがコンピュータを用い簡単に利用できることによる能率の向上は少なくないだろう。すでに色のレイアウトなどに関しては、形容詞情報を用いて情報を検索し、パネル作成を行なう支援システムも紹介されている[Kasahara 1988]。心理学を応用してデザイン支援に関する研究領域は感性工学として位置づけられている[Nagamachi 1990]。もう一つは、知的インターフェースあるいは人間型のロボットとして位置づけられる応用システムである。その中心的な課題は先程にも述べたように人間流の五感の作成である。将来的には人間の表情を読み取って、気を利かすというような人工的秘書と言うような具体的なアプリケーションの期待される領域である。

以下に感性情報処理の対象となる応用分野をまとめる。

1. 芸術における応用

例えば、音楽情報処理システム。これについては本論文の主題となっている。

2. さまざまなデザイン支援システム

デザインで使われる具体的な情報、例えば色とか形などを形容詞空間で記述するデータベースを作り、データの検索、変更などに利用する。さらにデザインに対する評価、アドバイスを行なう。

3. メディアの変換

絵画にあった音楽を選択するシステム、自動翻訳電話（どのような訳語を選ぶか、あるいはどのようなイントネーションが適切かを判断できるもの）など

4. マーケティング

絵画にあった音楽を選択するシステム、自動翻訳電話（どのような訳語を選ぶか、あるいはどのようなイントネーションが適切かを判断できるもの）など

4. マーケティング

従来のデータ分析法に加え、そのイメージと商品の具体的な特徴との関係に具体的な指標を与えることにより、商品の開発、および販売戦略に対するコンサルティングを行なう。

5. 知的情報通信

伝達したい情報を極度に圧縮し、受信側で再構成することによりデータを送信するという形態。例えば、テレビ電話などにおいて、顔画像を全て传送するかわりに表情情報をシンボル化して送信するなどの研究がある。

1-3 本研究の意義と論文の構成

従来の音楽情報処理システムは音楽という芸術分野を扱っているにもかかわらず、他の情報処理と同様、処理の対象となるのは数値的な情報や客観的な知識であった。これに対し、音楽感性情報処理は、今まで積極的に取り上げられなかった音楽情報の感性的側面を取り入れた、すなわち、音楽情報処理と感性情報処理の融合した研究領域である。前節までに述べてきた音楽情報処理、感性情報処理の背景が、音楽感性情報処理の、科学的・工学的意義、さらに社会的意義につながっていることは言うまでもない。そして、このような位置づけの下に行われる本研究成果は今後の情報処理研究に一つの方向を与えることが期待される。

本研究は、音楽というある意味で非常に主観的な対象、しかも感性という本来人間の心の中で扱われてきたものを含めて考察を行うものである。このような点からは、本研究で扱う対象は、認知科学の扱う領域と一致するところもあるが、主題としては、人間の「心」の科学、あるいは「知」の解明というよりも、どのようにすれば感性情報をコンピュータ上で扱うことができるかという点においている。言い換えると音楽という題材を通じて、感性情報の工学的な扱いを確立することが本研究の最も大きな主題となっている。その一方で、実システムの構築を通じてモデルを検証するというここでの立場から認知科学に近い章や、また、人工知能としての色彩の強い章もある。

本論文は、7章からなり、主題は、第2章から第6章に述べる。第2章は、音楽感性情報処理システムKANSEIの概要について述べる。第3章と第4章までは音楽聴取過程、

第5章は音楽解釈過程、第6章では作編曲過程を扱っている。

第2章では、まず、音楽感性情報処理の中心的課題を述べる。そして、音楽聴取という点を中心に、音楽分析、理解における基礎的研究を紹介したうえで、音楽感性情報処理システムKANSEIの概要、低次音楽情報処理、音楽感性情報処理を紹介する。

第3章では、採譜処理に関する諸問題について述べ、その基礎となる周波数解析法について述べる。次に、多重音の採譜システムについて述べる。これは、音楽理解のデータ受容過程に相当するものである。さらに、音のモデル、タスクモデルに基づいた複数楽器による楽音の音源分離モデルについて述べる。

第4章では、音楽聴取過程の中でも高次処理に相当するもの、すなわち、音楽認識・理解について考察を行う。認識の結果はメロディ・コード進行などの分析だけではなく、感想文の出力という形で行なっている。さらに、この章では感性情報の記述、音の予測モデルなどについて述べる。

第5章では、音楽解釈過程、すなわち、演奏者が楽曲を自分なりに解釈を行って表現を行う過程のコンピュータ上での実現を扱う。章の前半では、演奏の生成過程として楽譜の認識：音楽構造の解析・表意記号の表現・音楽構造の表現について述べる。後半では実際の演奏からの演奏法の修得について述べる。楽譜情報を利用した効率的な演奏表現の抽出、表意記号、モチーフ、フレーズなどの表現ルールの発見、モチーフ抽出の戦略の学習などについて述べる。

第6章では、感性やエモーションを意志決定の制約条件に用いた作編曲システムについて述べる。従来の作曲システムでは意志決定法には乱数が用いられてきたが、ここでは音の予測性と意外性を結び付けた意志決定法について提案を行う。

最後に、結論となる第7章では、音楽という対象にデータ変換処理、「心」の実現の構築を目標としてきたKANSEIシステムの大局的な評価を行い、音楽情報処理、感性情報処理の今後の行方に対し考察を行う。

第2章 音楽感性情報処理

音楽は芸術の中でも早くからコンピュータの利用が進み、自動演奏、楽譜印刷、採譜、作曲及びその支援などのような処理を総称した音楽情報処理という分野が成立するに至っている。しかしながら、音楽情報処理自体が音楽製作のツールとして発展してきたことから、それぞれの領域での研究は個別に進められたものであり、人間の知的処理に相当するものとして総合的に音楽を取り上げた研究はほとんどなされていなかった。特に音楽情報の感性的側面は本来芸術分野において最も重要であるにもかかわらず、さき送りにされてきた領域である。本章では、音楽感性情報処理の課題について概観し、具体的に音楽感性情報処理を実現したシステム“KANSEI”の概要について述べる。

2-1 音楽感性情報処理

2-1-1 音楽の理解

音楽感性情報処理の目的を一言でいうと、音楽がわかるコンピュータシステムを構築することになる。

音楽がわかるということに関して最も直感的に思い付くのは“聞く”すなわち音楽聴取である。その他に作曲したり、解釈したりすることも音楽がわかることの現われであると考えることが出来るが、ここでは、まず、聴取過程を対象に工学的な立場からの音楽の認識、理解について考えて行くこととする。

音楽聴取にみた認識・理解の過程はパターン認識で取り組まれてきた過程に最も近い過程と考えることが出来る。画像理解、音声理解、音楽理解の対比を図2-1に示す。

近年のパターン認識の研究においてはモデル駆動やトップダウンというような手法が多く見られるが、図2-1においては、理解に対する概念的な関係を示すために、対象データから内部表現に向かう処理、すなわち、ボトムアップ的に記述している。

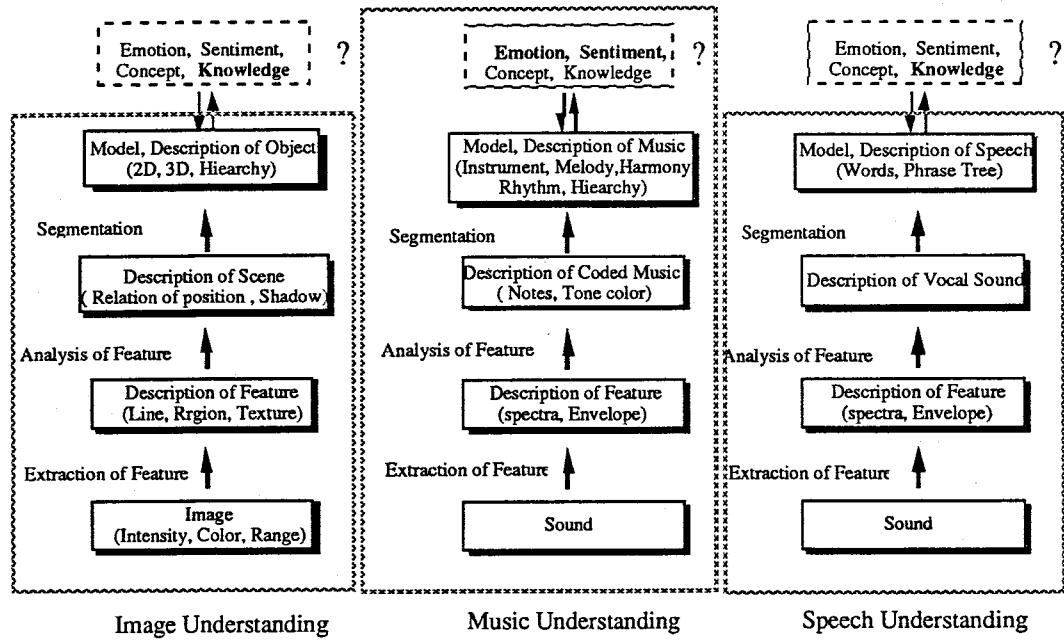


図2-1 画像理解、音声理解、音楽理解の対比

音楽聴取の根底にくるべきものは音響信号の音楽的知覚である。これに相当する具体的研究には、周波数分析、音源分離、拍の割当、音源定位などを挙げることができる [Katayose 1990 a]、これらはシステムの構築という立場からは自動採譜という領域で取り組まれてきたものである。図2-1においては、特徴の抽出、特徴の解釈までが、音楽の知覚に相当する。しかし、この段階では、まだ音楽としての認識・理解は行なわれていない。採譜処理そのものは音楽認識の入力段に相当するものである。

音楽における認識とは音列としてのシンボルを解析して音楽的な構造的特徴を記述することに相当するものと考えられる。例えば、メロディを抽出したり、コード進行を解析したり、ある音列の楽節における機能を分析・記述することが音楽認識となる。楽譜には音響を解析して拍子や音部といった構造的特徴が盛り込まれる必要がある。このような点からは、楽譜を生成するにまでいったシステムは一種の音楽認識を実現していると考えることが出来る。工学的なアプローチにおいても音楽の理解に対する考え方がある。いわゆる採譜システムを音楽理解とする人もいれば、音楽の認識に至れば音楽理解とする立場もある。後者の考え方はパターン認識の主要なテーマとなっている画像理解、あるいは、音声理解に対する一般的な見解と一致するものである [Kitahashi 1988]が、本研究においては、音楽理解とは音楽認識からさらに何らかの形で

心理量にまで至ることであるという立場を取るべきだと考えている。芸術の伝えたい情報は客観的情報ではなく、それによって想起される主観的情報であると考えるからである。もちろん音楽によって想起される感性のすべてがコンピュータで扱える形態であるはずはないし、仮に機械上で表現出来たとしても音楽のシンタックスとセマンティクスの関係に対して精密な理論を立てることは非常に困難となろう。また、音楽におけるサブシンボリック情報やメタファは元来言葉で表すことすら出来るものではないという議論もある。しかし機械が「あるメロディが基音から短三度はなれて音が使われている」という客観的記述より「短調で悲しげな曲である」という認識を持っているほうが、より人間的であり、自然なインターフェースが実現されていることは誰もが認めることであろう。ここでいう音楽理解の実現は実験心理学の蓄積してきたデータと方法論とコンピュータ技術の融合のうえに考えるものである。例えば、実験心理学においては調性の与える影響[Nakamura 1989]や演奏者の表現がいかに聴取者に受け入れられるかについてなどの研究が行なわれており、心理量が形容詞で測定されている[Senju 1988]。物理的な要素の機械が認識する技術を確立すれば、心理学が蓄積してきた膨大なデータの利用が期待される。

以上、工学的な観点から大雑把に音楽理解について眺めてきたが、音楽的構造の記述にいたる過程について考えてみる。これらの多くは現在、音楽認知という領域で積極的に取り組まれている領域である。その最も基礎となるものはオーストリアの音楽学者 Schenker (1867～1935)の理論であろう。Schenkerが理論を作った当時は認知科学やコンピュータといったことを意識しなかっただろうが、客観的な観点から階層的に解析を行なって音楽の骨格構造を抽出するという方法は現在の音楽AIシステムの研究に大きな影響を与え続けている。現在、Schenkerの理論はLisp風の記法で実現されたり、また、和声付けという具体的なアプリケーションにも応用されている[Smolian 1980, Ebcioğlu 1984]。Schenkerの理論が多分に和声的な理論から解析を行なうのに対し、聞き手の認知という観点から音楽のグルーピングを試みたのがMeyerである[Meyer 1956]。Meyerは「一つのリズム・グループが一つのアクセントしかもたない」という原則と音のまとまり方に関する心理学的知覚ルールを用いた音楽解析法を示した。この方法の長所は音楽の構造理解そのものに認知的要素を考慮したことである。が、計算理論という観点からはこの長所自身が欠点となっている。すなわち、この方法の根拠とするアクセントの決定論自身に音楽センスともいるべき主観的要素が要求されるために、結局のところ数理的に構造解析法を示したというまでには至っていない。また、Meyerは当初小節という単位をアクセントの解析ひいては構造解析のよりどころとしていたという問題もあって、後に

弟子のNarmourとともに上記の構造解析法にかわり、音楽進行に於ける局所的な期待感のボトムアップ的な重ね合わせが音楽構造を作るという構造解析法を発展させている。機械的実現という観点からは非常に煩雑な手続きを踏まねばならないが、人間的にという意味で音楽がわかるコンピュータを作っていくときには考慮して行くべき理論であるといわれている。

音楽感性情報処理の中心的な部分は、音楽認知あるいは音楽学や心理学の領域で議論され続けてきた領域である。それらの研究を参照し、利用して行くことは非常に重要なことである。しかし、これらの研究は人間の音楽プロセスに対するより深い理解を眼目としており、それ自体コンピュータですぐに使えるアルゴリズムを提供するものではない。また、人間であれば当然という範囲で理解できることが実は非常に難しい問題を含んでいるため、コンピューテーションナル・モデルを作るということ自体を音楽認知科学の大きな研究テーマの一つとして取り上げていくべきである。

計算科学的なアプローチで、最も初期のものに属するのがWinogradの調性解析である[Winograd 1968]。ここでは、和音の記号列の入力に対し生成文法的アプローチで解析を行ない、調性およびその調性におけるコードの機能が解析される。伴奏を伴わない旋律のフレーズ解析としては、ゲシュタルト心理学のヒューリスティック・ガイドラインを応用したものも発表されている[Tenny 1980]。現在、音楽認知の分野でもっとも関心を集めているものの一つがLerdahlとJackendoffによって提唱された階層論である[Lerdahl 1983]。この理論ではシェンカーフ音楽理論とチョムスキーフの文法を概念的に結合した音楽の構造解析法が示されている。それ自体は具体的なアルゴリズムを示しているわけではないが、音楽の文法的ルールに対する厳密な考察から、アプリケーションシステム、特に、自動作曲や編曲支援システムにおいて、開発者の意識するしないは別として、Lerdahl流の考え方を受け入れられてきた。これに対し、批判の声がないわけでもない。音楽を言語のように扱ってはならないという議論である。言語であれば、単語というシンボル情報が文全体の意味を決定する最小単位となるが、音楽においては、音楽を構成する最小単位の要素が明確に決定できない。仮に、最小単位の単語に相当する音のまとまりが与えられたところで、セマンティックが定義できないというわけである。このようなことから、シンボリックな表現ではなくて、サブシンボリックな音楽表現をどのように実現すればよいかというパラダイムが持ち上がっている。現在、解決を見たわけではないが、一つの方法論として、並列処理やコネクショニズムによるアプローチが現在の流行となっている。リズムの認識、作曲など、それぞれの対象とする領域に応じた並列の処理系の実現が試みられている。

また、計算科学的アプローチにおいては対象の記述法を確立することが理論の構築や具体的なアルゴリズムの作成において非常に重要である。平賀は音楽を関数形式として抽象化することにより知識を抽象化された音楽表現に対する操作として記述する枠組みを与えた[Hiraga 1989]。音楽知覚認知の解明という抽象的な問題が音楽表現の操作の戦略を探るという具体的方法論の構築に置き換えることができるため、音楽認知の新たな発展に寄与するものとなろう。

2-1-2 作曲プログラム

前節では、音楽の理解を工学的アプローチ、音楽学、そして認知心理学等で取り組まれてきた研究を紹介するという形で概観してきた。本節では、AIと音楽との関連領域として最も早くから取り組まれてきた作曲プログラムを歴史的に振り返り、音楽感性情報処理の意味合いを考えてみることにする。

作曲作業のための数学的な公式（ルール）の作成はルネッサンス後期から始められており、正しい用例と誤例を示した集積的な楽典は1950年後半に作られている。このような“経験的な研究”に対してコンピュータとAIは、どのように作曲家が考えているのかを扱う定量的モデルの形成し、またモデルの評価に対する非常に有望なツールとなりえたため、いわゆる作曲プログラムの研究が行われるようになった。ここでいう作曲プログラムはすでに存在している作曲法のモデルを検証するための解析ツールと新しい芸術作品を作るための創作的ツールとしての意味合いを持っている。

HillerとIsaacsonの弦楽四重奏のためのイリアック組曲における作曲プログラムにおいては18世紀の対位法から導きだされた簡単なメロディとハーモニーのルールを用いて音高の決定を行なっている[Hiller 1959]。メロディ生成時に使われるルールは、一つの音高を3回もしくはそれ以上繰り返したり、連続音のあとにはスキップ音をいれるというようなものである。ハーモニーのルールは2度の音と7度の音のような不協和音に制限を加えるものや平行完全協和音でのモーションを禁止するものや対位のモーションを要求するものなどである。意志決定機構は乱数要素により、ルールによって定義された“正しい”音楽世界の範囲内で均等的な音楽ルールの選択を行なうという形で実現されていた。

乱数要素を取り入れた意志決定法は、以降の作曲や編曲システムにおいて、音高にとどまらず、リズムや音色など他の音楽の属性に対しても広く用いられるようになった。

しかし、80年代になって、例外に対する処理とすべてのルールにマッチしない属性の存在に対する問題点が指摘されるようになった。これらの問題に対する方策として、絶対的な基準をルールとして記述し、それ以上の基準はヒューリスティック・ガイドラインを変更することによって例外処理を行なうという方法がとられるようになった[Ebcioglu 1984, Ames 1983]。ヒューリスティックな手続きとしては、他の競合選択で最も使用される頻度が少なかった音高に優先度を与えることによって音高の統計学的なバランスを維持するというような機構が用いられる。2番目の問題点に対する対策としては、バックトラック機能が用いられている。ある時点に於ける音高を決定するためのルールが一つも見つからなかったときに、その矛盾を引き起こす原因となった前の音を修正するという作業が実現されている。

楽曲の階層性を意識した研究においては、自然言語処理に於ける技術を応用した研究が、70年代後半から見られるようになった。その例としては、Holtzmanによって開発された文脈依存文法を用いた音楽の表現などがある[Holtzman 1980]。音楽知識を直接プロダクション機構に組み入れたシステムも提案されている[Ames 1982]。また、ATNやオブジェクト指向を組み入れた作曲支援環境も提案されている[Cope 1987]。これに対し、音楽が他の知識情報処理に比べ、知識という単位での音素の扱いは不可能であるという議論もある。特に、自然言語のようにワードの単位で分解されたいわゆるシンボル表現をそのまま音楽に適用することに対する問題点が指摘されているが、実用という観点から成功しているシステムにおいては多かれ少なかれシンボル的アプローチをとっている。

感性情報の扱いの必要性が特に意識されるようになったのは1980年の後半になってからである。先駆的な研究としては、音の大きさや音色等の主観的な音楽の属性とアンプリチュードや周波数といった客観量の対応関係を決定するために心理学データを解析するといったものもある[Martens 1985]。作曲など高次のレベルの処理に感性情報を考慮したシステムは1989年に発表されている[Riecken 1989, Saiwaki 1989]。前者はMinskyが提唱したK-lineを用いた形容詞ベースの作曲システムとして注目を浴びた。なお、後者の形容詞ベースの作・編曲支援システムについては第6章で述べる。

自動作曲プログラムに関する研究の歩みは感性情報処理に対して非常に興味深い示唆を与えている。初期においてはとにかく外側から見て知的な働きをするシステムの構築が進められた。が、その段階でシステム自身がなぜそのような処理が行なわれているかについてはシステムが理解するものではなかった。これはどのようなルールが使われたかを決定する意志決定にランダム信号が使われている点に端的に現われている。このような点に対する反省もあって、例外処理とそうでない部分を明示的にプログラム化した

り、矛盾を回避するための意志決定を前面に押し出した処理方式が組まれるようになった。このことは、機械に処理に対する“理解”を実現させようという努力にはかならないものである。同様のことは、言語理論の応用した作曲システムの研究にも言えることである。そして、現在、感性情報を考慮したシステムの研究が行なわれつつある。これは処理に対する“理解”ということから一步踏み込んで、その処理がなされた理由を機械に“理解”させる方向に興味がむいていると考へても間違いではないだろう。

知識工学のなかでも中心的に研究されてきたエキスパートシステム、CAIにおいてもシステムが処理に対する“理由”を理解している必要があると言われている。音楽は特に感性と密接に結びついた領域ではあるが、作曲プログラムに見られるこの流れは、他の知的システムにおいても進んで行くものと思われる。

2-2 音楽感性情報処理システム“KANSEI”

前節では、音楽感性情報処理の目的は音楽がわかるコンピュータを作るということであると述べた。それと同様に重要なことは、人間が普段音楽にかかわっているのと同じ形で外界情報をやり取りする機構を実現することである。AIの分野で処理の対象となってきたのは、音列情報、すなわち、キーボードでやり取りされるシンボル情報であった。しかし、人間は音楽を音響あるいは楽譜など聴覚・視覚を介する情報として知覚し、また、生成を行なっている。もちろんシンボルレベルで扱われる音列情報も感性を担う重要な要素であるが、情緒溢れる演奏や非常にきらびやかな音など、音楽における感性的な情報は今まで情報処理において捨てられてきた音響信号におけるパターンに起因することが多い。また、楽譜にしてもその構造的な美しさというものは楽譜中のシンボルレベルではなく画像としての均齊等に大きく依存するものと考えられる。したがって、音楽感性情報処理システムを作つてゆくときには、システムの心の部分を作るとともに、音響を知覚する耳、楽譜を読む目、そして演奏を行なう手足、口等の機能を実現することが不可欠である。

“KANSEI”は、聴覚・視覚にわたるデータ入出力機構、音楽理解能力を有する総合的な音楽感性情報処理システムである[[Katayose 1989a, Katayose 1990d]。このシステムの概要を図2-2にハードウェア構成を図2-3に示す。

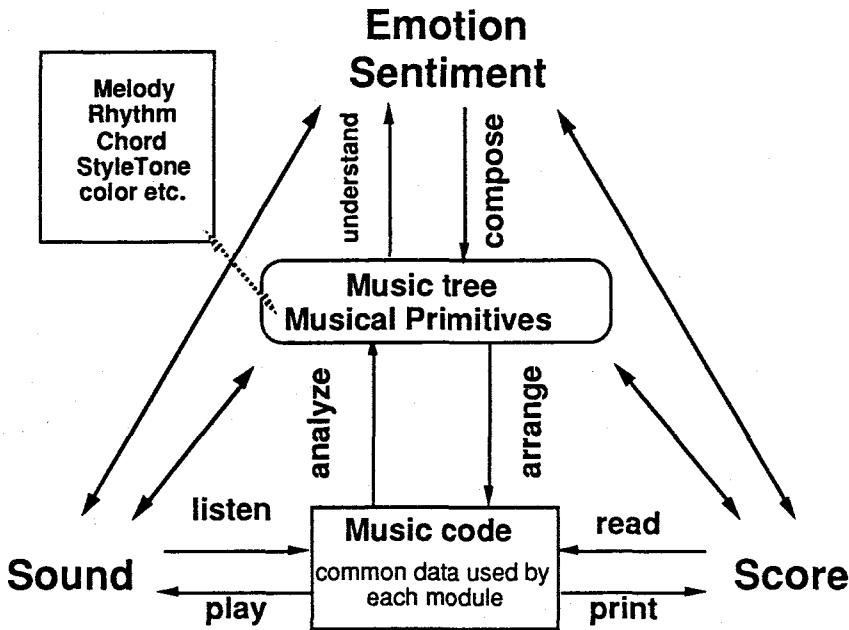


図2-2 感性音楽情報処理システムの概要

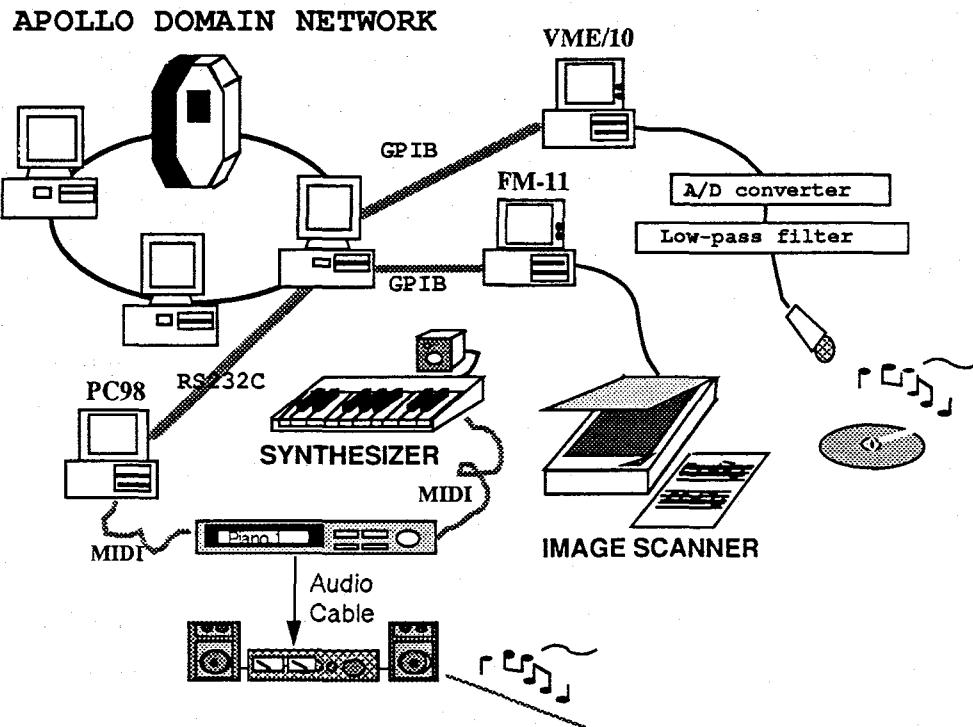


図2-3 ハードウェア構成

図2-2において、Music Codeとは音列を表すシンボルを意味している。Musical Primitive（音楽プリミティブ）はメロディ、リズム、曲の形式などの音楽的要素をさしている。

2-2-1 低次音楽情報処理過程

本システムのデータ入出力機構を以下に示す。

(1) 採譜[Katayose 1991]

音楽音響信号からシンボル情報に変換する。本システムでは対象の音楽制約や知識を用いることにより、メロディだけではなく、自動採譜システムの対象とされていなかつた多重音の採譜が可能となっている。通常のピアノ曲に対しては認識率は90%を越えている。現在、複数の楽器による楽音の採譜も手掛けている。

(2) 演奏特徴抽出[Takami 1989b]

名演奏における情緒を理解するためには、演奏の正確な同定が必要となる。本システムにおいては、楽譜情報を利用することにより正確な演奏情報の抽出が可能となっている。この処理は譜面を思い浮かべながら音楽を聴く過程に相当する。

また、完全な楽譜化にまでは至らないが、尺八を代表とする和洋伝承楽器の楽音を分析することができる。

(3) 演奏

本システムにおいては、MIDIを用いて表情豊かな演奏を行うことが可能である。現段階では16チャンネルで音の始点・終点・強さのコントロールができる。現在、知的処理部はワークステーション上に実現されている。それ自体は自動伴奏のようなりアルタイム処理はできないが、パケット的に演奏データを通信し、演奏を行なうことができる。リアルタイム処理はMIDIのドライバとして使用しているPC9801において実現されている。

(4) 読譜[Kato 1988]

印刷ピアノ楽譜の読譜が可能である。楽譜に記載される各々のシンボルの形状は比較的単純であるが、実際の楽譜においては、シンボルの交差、接触が頻繁に起こっている。

また個々の記号がもつ意味や連結関係にはあいまいさが存在する。本システムでは、小節単位に処理を分割し、意味解析においては、一種の黒板法を用いることにより、上記に問題に対処している。通常のピアノ曲に対しては認識率は90%を越えている。

(5) 写譜[Katayose 1990a]

音楽情報処理システムにおいては、内部データを人間にわかりやすい形で出力することが重要である。視覚的な情報出力としては楽譜化が望まれる。楽譜化においては、和音、休符、タイなどの音楽的解析が必要である。本システムでは、このような音楽的解析を行ったうえ、階層的なグラフィックデータとして出力することができる。また、そのハードコピーが利用できる。

2-2-2 音楽感性情報処理

1. 音楽聴取処理[Katayose 1988b]

1) コード進行、調性、メロディの認識等

採譜、読譜等の処理により入力した音列シンボルから、音楽理解の土台となるコード進行、調性、リズムなどの解析を行なうことができる。

2) 感想文の出力、聴取モデル

上記の認識結果（音楽プリミティブ）とそれに対応する形容詞のリンクから感想文が生成できる。これは最も簡単化した感性情報処理の形態であると考えられる。また、従来から心理学実験で用いられてきた方法に基づいて感性量の次元に音楽プリミティブと形容詞を対応させ、また、時間関数と次の音列、進行を予測するモデル等を用いて疑似的な鑑賞モデルが実現されている[Katayose 1989 c].

2. 音楽解釈処理

1) 演奏ルールの適用[Katayose 1989]

読み取った楽譜に対しモチーフ、フレーズを解析したうえで、そのモチーフ、フレーズの表現と表意記号の表現を加えることにより、情緒溢れる演奏を生成することができ

る。また重々しく、弾んでといったような形容詞情報を用いて表現を強調することができる。これは演奏における物理的パラメータと音楽の構造を多義的に対応付けた感性情報処理の具体的システムとなっている。

2) 演奏ルールの抽出・学習[Katayose 1990 b]

実際の演奏から、音楽解釈を行うためのルールを抽出することができる。これは、感性情報の学習に相当するものである。

3. 作編曲およびその支援[Saiwaki 1989]

1) 感性パラメータの利用

形容詞情報で作曲・編曲を行なうときに必要な音楽プリミティブを選択することができる。その発展として、言葉による自動作曲、画像にあった楽曲の選択等を行なうことができる。

2) 予期モデルに基づいたメロディの生成

従来の作曲システムにおいては、ランダムネスが意志決定の重要な手段として使われてきた。“KANSEI”では、意志決定において、次の音列が予測可能かどうかという要素を意外性・期待度という感性的な要素と結び付けて、生成したメロディの評価を行なうことができる。

2-3 音楽データ

“KANSEI”においては複数のデータが使われている。データは音列データとコードやメロディなど音楽の構造的単位を表す音楽プリミティブに大別される。本節では、音列データについて述べる。音楽プリミティブについては第4章で詳しく述べる。

音列データと一言にいっても、楽譜に即したデータ、演奏に即したデータ、構造解析に即したデータなど様々な用途に対して使用できる必要がある。本研究では、すべての用途に対して使用できる一つの一般的データ形式を使用するのではなく、それぞれの用途に合った外部音列データを用意し、それぞれのデータを相互変換する機能を実現することによりシステムとしての統合を実現するという方針を取っている。各外部データの

相互変換を行うための内部データとしては、時間軸方向におけるリストに各音と属性情報をリストとして格納する方式を用いている。全体的なシステムの統合はそれぞれのデータの相互変換ツールを用意することにより取られている。外部データには楽譜情報に密接なscr形式、音列情報に密接なnote形式とMIDIデータの送信用のbin形式がある。bin形式はMIDIの規格に非常に近く、データ転送のない区間をランレンジスを用いてデータ圧縮を行なった形式である。scr形式、note形式については以下に紹介する。

2-3-1 scr形式

scr形式は、楽譜情報を表現するために使われるデータである。このデータの表現例を以下に示す。この表現では、各小節に於ける長さと分解能を宣言し、それにしたがって、各音のデータを記述するようにしている。また、このデータ形式では、属性記号とその使い方は各自の定義を許している。各シンボルの位置情報を書きたいときには属性として記述する。基本的な音素の書き方は以下に示すとおりである。

音名 スタート時間 音の長さ (属性記号 [そのタグ] , [属性記号 [そのタグ] ,])
個々の音にかかる表現記号は、() 内に記述する。

[例]

C3 0 24 (right, slur 1, <)

一つの音名以外にかかる表現記号等には、以下の表現を用いる。

識別子 スタート時間 [タグ] (表現記号 [タグ])

[例]

!! 0 (pedal on)

このフォーマットを用いた楽譜データの例を以下に示す。

```

composer F.Chopin      <- 作曲者
songname Walzer        <- 曲名
key Eb                  <- 調性
Tempo quarter 0.2 (秒) または Tempo giusto <- テンポ
==<1>== 24 * 3 <- (4分音符のDIVISION * 拍子の数、デフォルトは 24 * 4)

```

```
!! 0 (pedal 1, piano)
C3 0 24 (right, slur_start, 1beam 1)
Ab3 0 6 (right, beam1)
Ab2 0 24 (left, beam 1)
Eb2 6 6 (right, cres 1)
Ab3 12 6 (right, slur_end 1, cres 1)
Eb2 18 6 (right, decre 1)
!! 18 (pedal 0)
Bb3 24 24 (right, decre 1)
G2 24 6 (right)
Db2 24 24 (left, cre 2)
Eb2 30 6 (right, cre 2)
G2 36 6 (right)
Eb2 42 6 (right, tie 1)
==>=
.
.
.
```

2 – 3 – 2 note形式

この形式は、表情を付けた演奏を行なうため、各音のスタート時刻のずれ、持続時間のずれ、テンポの変更、休止区間の挿入、音源の変更などを明示的に記述できる形式となっている。音の開始時刻は絶対時刻で記述する。

通常の表記では、4分音符の長さを1として
 スタート時間 (音名 長さ) (音名 長さ) . . .
 の形式を取る

[例]

```
Header{
BEAT 4/4
TEMPO 0.8
MIN_UNIT 1/4
}
/* 弱起の時はここに書いててもよい */
----- 0 /* 小節線の位置 書かなくてもよい */
0 (G3 1) (A2 2) (D2 0.5)
0.5 (F3 0.5)
2 (B3 1)
3 (F4 2)
```

```
----- 1 /* 小節線の位置 */
```

```
.
```

```
.
```

```
END
```

演奏に変化を付けたいときは以下の形式を用いる。

スタート時刻 (音名 スタート時刻のずれ 長さ 長さのずれ ベロシティ (音の大きさ)) . . .

[例]

```
Header{
```

```
Beat 4/4
```

```
EXP ON
```

```
Tempo 0.5<---- 4分音符のテンポ
```

```
}
```

```
/* 弱起の時はここに書いててもよい */
```

```
----- 0 /* 小節線の位置 書かなくてもよい */
```

```
0 (G3 +0.0 1.0 -0.1 60) (+0.1 A2 1.0 -0.1 46) (+0 D2 0.5 -0.05 52)
```

```
0.5 (-0.1 F3 0.5 +0 24)
```

```
2 (+0.2 B3 1 +0 23)
```

```
Tempo 0.5 <--- テンポの変更
```

```
3 (+0.05 F4 2 +0 78) (D3 +0.1 2.0 -0.1 45)
```

```
Rest 0.2 <--- 休止区間の挿入
```

```
----- 1 /* 小節線の位置 */
```

```
.
```

```
.
```

```
.
```

```
END
```

2-3-3 内部データ

内部データは、リストに基づいており、各データの変換等を行なうときに使われる。内部データの概念図を図2-4に示す。これは、作曲支援環境などで使われているデータ表現と類似のものである。図において、各ノードは、待ち時間（ウェイト）で結合されている。ノード、ノートに対しては、属性情報（表意記号やペダルなど）はアトリビュートをリストとして付加することが表現する。

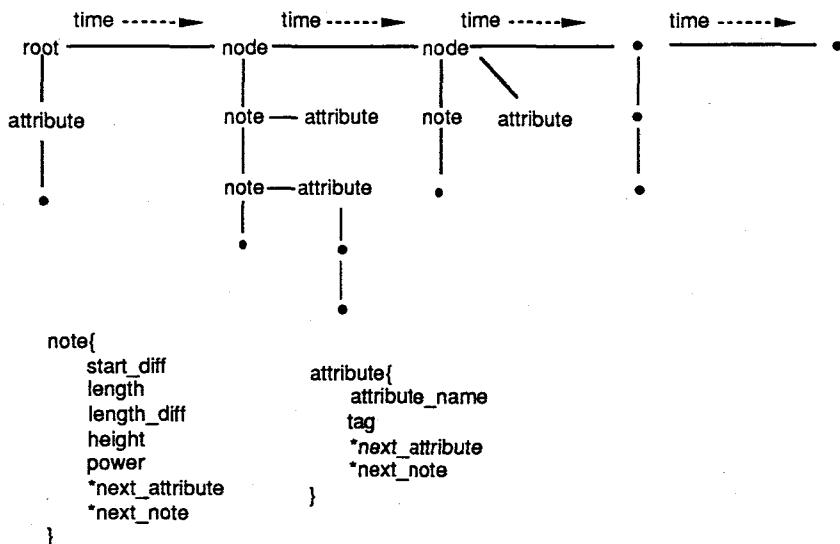


図2-4 内部表現

この形式において；ノードを一つのプロセスとし，時間の制御をプロセスの機動・ウエイトなどの処理するデータ処理部を実現すれば，リアルタイム処理にも非常に適したものとなっている。

音楽認知・知覚を考えるときには，その構造を記述することが望まれるが，ここでは属性情報に構造を表すタグを付加することによって構造情報を記述することができる。

2-4 まとめ

音楽感性情報処理システムは従来のデータハンドリングを主眼とした音楽情報処理システムに対し，人間に見られる音響・画像にわたるデータ入出力機構，音楽分析・理解の実現を目指した総合的な音楽システムである。

データ変換機能を以下にまとめる。

- ・採譜
- ・演奏特徴抽出
- ・演奏
- ・読譜
- ・写譜

高次音楽情報処理を以下にまとめる。

(1) 音楽聴取処理

- ・コード進行、調性、メロディの認識等
- ・センチメントの抽出、観賞モデル

(2) 音楽解釈処理

- ・演奏ルールの抽出・学習
- ・演奏ルールの適用
- ・モチーフ、フレーズの認識と類似度の判定
- ・音楽概念の表現

(3) 作編曲およびその支援

- ・感性パラメータの利用
- ・予期モデルに基づいたメロディの生成

(4) メディアの変換

- ・言語からの作曲
- ・音楽に合った画像の選択

第3章 自動採譜

この章では単一楽器によるポリフォニーを採譜するシステムおよび異種複数楽器による音楽の採譜機構について述べる。従来の自動採譜システム共通の問題としては、処理対象に対する制限の厳しさ、柔軟性のなさをあげることができる。これは人間の専門家にみられる知識処理が欠如しているからである。まず、この観点からの考察を加えることにより作成した単一楽器によるポリフォニーの自動採譜システムについて述べる。このシステムは、音楽感性情報処理システムの中核として、データベースの構築、センチメントの抽出の情報知覚部として実働している。実験としては、ピアノ、ギター、三味線を対象楽器として採譜処理を実現し、本システムの有効性と限界を検討している。次に異種複数楽器による音楽の採譜機構について述べる。ここでは、ハーモニックエンベロープに代表される音のモデルとタスクのモデルを持ち合わせたうえで、最適な認識法を自動的に実現する音の知覚モデルを提案し、そのモデルを用いた基礎的な実験について述べる。

3-1 自動採譜における諸問題

音響からのデータ入力は音楽を扱う上で最も根本的な過程である。この過程を表すのに通常“採譜”と言う言葉が用いられてきた。文字どおりに“採譜”を読むと「譜を探る」、すなわち、音響から譜（楽譜）を作るという意味になる。また音楽教育の分野では、ピアノなどの和音の名前を当てることを“採譜”と呼んだりしている。信号処理の分野においても採譜とは、単に音響からスペクトログラムを抽出するものから、楽譜を出力するものまでその使用範囲は多岐にわたっているが、ここでは、音響を入力として楽譜を出力する過程を広義の採譜、音響から音素をシンボルとして抽出する過程を狭義の採譜と定義する。以下、特にことわりがない場合は、自動採譜を広義の意味で用いる。

自動採譜の難しさは対象によって大きく変わるが、一般に処理すべき諸問題を以下にあげる。本来、自動採譜は音響工学の一対象として研究されてきたものであるが、人間の処理の代行という立場からは知識処理と扱うべき点も多い点を注意されたい。

3-1-1 音響信号から音素シンボルへの変換過程

この過程は、ほとんど意識に止めることなく行われる知覚過程である。ここで使われる知識は明確な論理的知識というより、音色、フレーズといったモデルとその活用法である。この過程の認識対象は、音の分離、拍の認識である。

1) 高分解能のピーク周波数の抽出

従来より周波数解析としてはFFT、自己相関関数などが用いられてきた。しかし、実用レベルでは、これらは方法は低音部においては、十分な分解能が得られないという大きな問題があったため、通常のシステムにおいては単音のメロディというような厳しい条件が課されていた[Kitagaki 1977]。このような問題に対する解決法としては信号処理の能力向上が不可欠である。これに対しては複素スペクトルの位相成分に注目した複素スペクトルの内挿法[Hara 1983]や各周波数における分解能が等しくなるように窓関数に工夫を加えたフーリエ変換Wavelet Transform [Kronland-Martinet 1987]などの方法が提案されている。本論文における周波数解析には複素スペクトルの内挿法を用いている。複素スペクトルの内挿法については次節で紹介する。

2) 基本波の抽出

楽器音には基本波のほかに高調波が存在するため、和音に含まれる周波数成分は非常に複雑なものとなっている。高調波はポリフォニーが対象となるとき音名を決定するのに非常に厄介な問題となる。たとえば、中点のCの音とそのオクターブ上のCが周波数解析によって抽出された場合、オクターブ上のCが実際に弾かれた音かどうかを判断するには、音色のテーブルおよび音数などのモデル駆動の処理を実現する必要がある。

従来の採譜に関する研究においては主に単旋律を対象としたものが報告されている[Niihara 1984, Fujii 1989]。多声を扱った初期の優れた研究としてはMoorerの研究がある[Moorer 1975]。しかし、これも対象を2声のギターに限定したものであり、一般多声を扱うものではない。Tanguianeは各周波数成分の（パワーではなく）オン・オフに注目した方法を提案した[Tanguiane 1989]。ある和音が鳴っているときの基本波および高調波のパターンを $s(i)$ とする。ここで、 i は半音階を単位とし、音階 i においてパワーが存在するときに $s(i) = 1$ 、存在しないときに $s(i) = 0$ となる関数である。連続するコードにおける

パターン $s_1(i)$ と $s_2(i)$ の相関関数

$$K_{1,2}(i) = \sum_{n=\max(l, i)}^{\min(N, N+i)} s_1(n-i) \cdot s_2(n), i = -N, \dots, -1, 0, 1, \dots, N.$$

を計算し、変化した周波数構造パターンに注目することにより基本波を抽出する。この様子を図3-1に示す。この方法は変化に注目するということからロバストである半面、高調波のオン・オフを決定するためのThresholdの問題、対象に音の強弱があるときの処理などの問題がある。

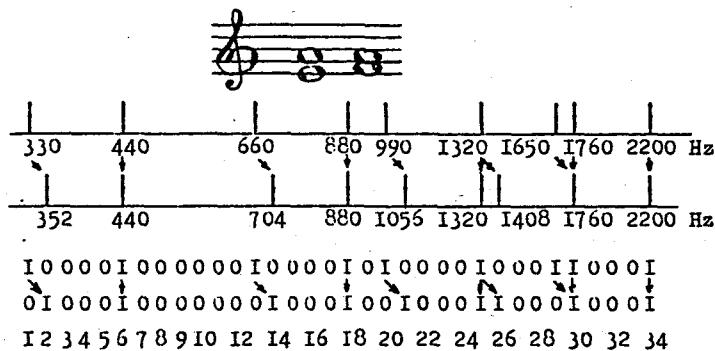


Fig.6.

Table I. The most salient patterns of partleading intervals between chords (eI, aI) and (fI, aI) for harmonic voices with 5 partials.

Salience of the pattern (correlation)	Size of the pattern in semitones	Numbers of frequency bands, which contain partials, generating the pattern of interval	Number of successive harmonics in low-level patterns	The pattern of interval in standard notation
6	-4	I8 I8 25 29 30 34	5	(aI, fI)
5	0	I8 I8 25 30 34	5	(aI, aI)
5	I	I I3 20 25 29	5	(eI, fI)
5	5	I I3 20 25 29	5	(eI, aI)
2	-7	I3 25	2	(e2, aI)

Table 2. The most salient patterns of harmonic intervals in the chord (eI, aI) for harmonic voices with 5 partials.

Salience of the pattern (correlation)	Size of the pattern in semitones	Numbers of frequency bands, which contain partials, generating the pattern of interval	Number of successive harmonics in low-level patterns	The pattern of interval in standard notation
9	0	I6 I3 I8 20 25 29 30 34	5	(eI, eI)
5	5	I I3 20 25 29	5	(eI, aI)
3	7	6 I3 I8	2	(aI, e2)

図3-1 相関関数を用いたコードの判別

3) 音源分離

上記の問題の最も一般的なケースとして、複数楽器による楽曲の音源分離がある。訓練を積んだ人なら、自分の聴きたいパートだけ聴くというような処理ができるが、これは、音色だけではなく、エンベロープ、ビブラート、音の定位などの情報を有効に用いているからできるものである。音源分離の研究としては、複数のマイクを用いて位置情報から音源分離を行うものが有名であるが、人間は一つのスピーカーからでてきた音楽に対して音源分離を行うことができる。これに対する研究は音のモデルを作成し、それに応じた処理を実現するという方法が取られる。本論文では、楽器の判別、2種の楽器音の分離という実例を取り上げ、音源分離を行なうモデルについて提案する。

4) ビート・トラッキングと拍の量子化

音楽に於いては、音高（音名）を同定すること以上に拍の割当が重要な問題となる。通常、人間の音楽聴取過程では音の立ち上がり時刻、テンポは拍、小節という単位で正規化したうえで認識される。ところが、実際の演奏においてはテンポを守って正確に演奏されることはほとんどなく、中にはひとつの音符の長さが楽譜上のそれの倍近くで演奏されることもある。音楽を記譜する際には、それぞれの拍の長さの比は有理数で表されるように、うまく量子化を行ってやる必要がある。この問題は、ビート・トラッキング、あるいは拍の量子化（Quantization）として認知科学としての感心も高く様々なアプローチから研究されてきた。Longuet-Higginsは、いわゆるロールピアノのロールにかかっている単旋律の音高、音の始点終点の情報を入力として、音価、調性の解析を行なうプログラムを作った[Longuet-higgins 1987]。ここでは、音楽のリズムは構成上、2分木または3分木により全体の音価を分解していく過程でできあがるものであるという前提を用いて、テンポに表情を付けた入力から、音符の階層的記述を導出している。最近では、Desain がコネクショニスト・モデルによる有望な方法を提案している[Desain 1989]。このアプローチでは、隣接する各音の間隔の長さの比が有理数に近づかせるようにする変化関数（図3-2）を作り、その変化関数の働きをネットワーク上の相互作用として実現し（図3-3），イタレーションを行うことにより最終的に拍の量子化を行っている（図3-4）。また、個々の音だけではなく、連続する二つの音の長さの和を一つのノードとしてネットワークの相互作用に組み込むことで、処理の向上が図れるとしている。これは音の知覚というレベルで非常に良い例であるが、音楽としての制約はあまり

利用していない。本論文では、実際の音響に基づいているため、ノイズ対策も兼ねて、拍の強さ、小節などの音楽的制約を利用し、次の拍を予想するという独特の方法を取っている。

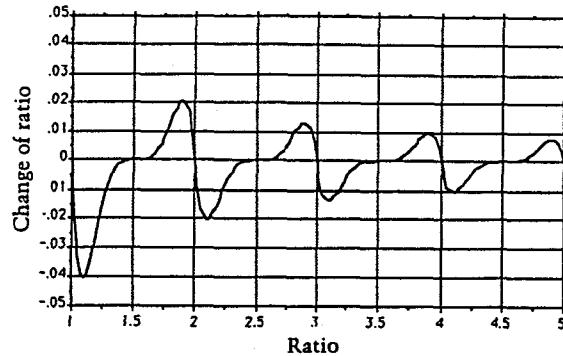
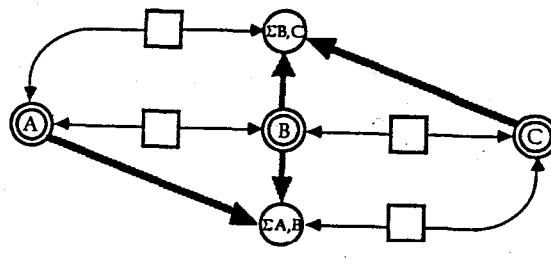
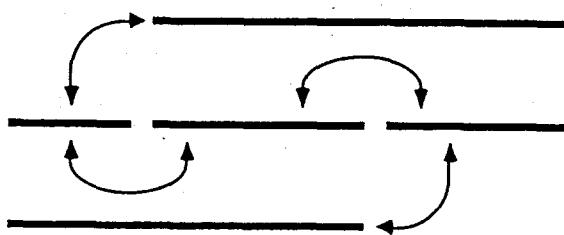


図3-2 変化関数



● Basic cell □ Interaction call ← Summation
 ○ Sum cell ← Interaction

a)

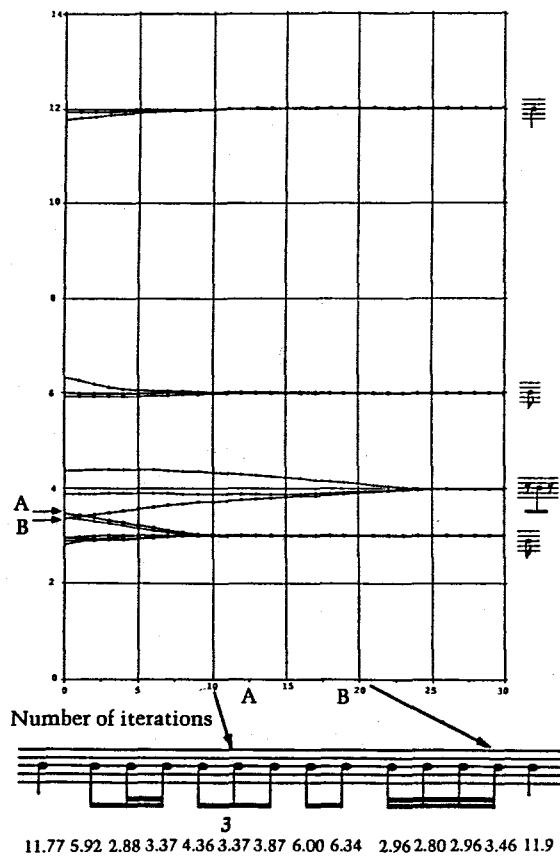


— Inter-onset interval

↔ Interaction

b)

図3-3 ネットワークによる相互作用の実現
a) ネットワーク b) 音の長さにかかる相互作用



a)



b)

図3-4 a) コネクショニストアプローチによるコンタタイプ
b) シンボリックアプローチによるコンタタイプ

3-1-2 音素シンボルから楽譜への変換過程

1) 構造解析

多重音を音響信号から解析して、楽譜を生成するところまでに至った採譜システムはほとんどないが、音列シンボルを入力とし、楽譜作成を支援するツールとしては実用化されているものもある。これらは美しい楽譜を生成することとデータ入力の簡略化を目指しており、レイアウトに関してはかなり知的なシステムも紹介されている[Oka 1987]。

美しい楽譜をつくることはもちろん重要なことであるが、音響レベルから自動採譜という目的からは人間が楽譜作成ツールを使ってシンボルを組み合わせる作業、すなわち音符（音素シンボル）の構造的把握が重要である。この作業は上記のように Longuet-Higgins の研究でも一部取り上げられているが、本研究では、楽譜生成にまで至っている。調性、拍子の解析、和音、連符などの楽譜にかかる情報は音列情報から直接的に得られるものではなく、シンタックスを解析して得られるものである。このほか、わかりやすい楽譜を生成するために休符を補う、段を分ける、タイを用いるなどの処理を行う必要がある。

2) 楽譜の音楽的補正

実際のデータでは、演奏の仕方により音のエンベロープが異なり、本来同じ長さの音符でも持続時間が長くなることとその逆に短くなることが混在している。人間の音楽専門家が採譜を行う場合には、特に意識もせずこれらの補正を行っているが、これには演奏法と音楽スタイルに関する一種の思い込みともいえる知識が利用されていると考えられる。この過程は、上記の音素レベルの処理と楽譜レベルの処理のどちらかに分類するより、処理全体を通じた知識処理と考えたほうが自然である。また、人間の歌唱の採譜においては、ピッチの補正が必要となることがある。この問題に対しては、調性を解析したうえで、その調にはずれた音を補正するという方法をとっている例もある[Niihara 1984]。

3-2 複素スペクトルの内挿法と音階の単位系

周期信号から周波数を抽出する方法としては、離散スペクトル解析（DFT）がよく知

られている。一般に離散スペクトル解析では、周波数の分解能は観測時間の逆数でおさえられるため、分解能をあげるために観測時間を長くする必要がある。ところが音楽音響信号においては、短時間で音高すなわち周波数が変化してしまうという問題がある。例えば、C3(130.81Hz)とC#3(138.59Hz)を分離同定するためには、少なくとも観測時間は

$$T = 1 / \Delta f = 260 \text{ [msec]}$$

でなければならないが、通常の音楽作品において一つの音が定常的に260 msec以上持続するとは限らない。このような性質上、音響信号からは、短い観測区間でかつ精度よく周波数を同定する必要があり、この問題を解決するためには、通常のDFTやFFTに対して何らかの改良を行ってやる必要がある。ここでは、複素スペクトルの位相特性に注し、効率よく周波数を解析する複素スペクトル内挿法の原理について紹介する。

3-2-1 複素スペクトルの内挿法

振幅A、周波数、観測時間、区間中央での位相 ϕ なる單一周波数波形の複素表現

$$x(t) = A \cdot \exp\left\{i \cdot \left(\frac{2\pi ft}{T} - \pi f + \phi\right)\right\}$$

を考えると、

この離散フーリエ変換 Z_m は

$$Z_m = \int_0^T x(t) \cdot \exp\left(-i \cdot \frac{2\pi mt}{T}\right) dt = \frac{A \sin \pi f}{(f-m)} \exp(i\phi)$$

となる。 $m=f$ 付近の様子を図3-5に示す。この図から以下の性質がわかる。

- 1) スペクトルの方向は全て等しく、その方向は、信号の位相 ϕ で決まる。
- 2) 方向成分 ϕ の逆数は m についての一次関数で表される。
- 3) 信号の周波数はその関数と周波数軸との交点で与えられる。すなわち、方向 ϕ における Z_m の逆数を ω_m とすると周波数 f は次式で与えられる。

$$\bar{f} = m + \frac{\omega_m}{\omega_{m+1} - \omega_m}$$

しかし、実際には量子過誤差、ノイズなどの影響があるため、 Z_m と Z_{m+1} の位相差は π にはならない。そこで、単位ベクトル u

$$u = \frac{(Z_{m+1} - Z_m)}{|Z_{m+1} - Z_m|}$$

を考え、 u と Z_m 、 Z_{m+1} の内積の逆数を用いて、周波数 f を次のように推定する。

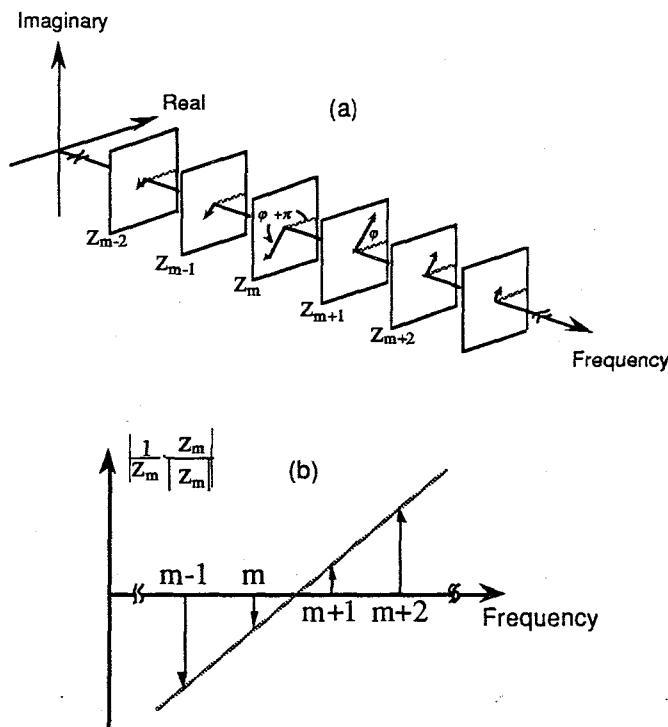


図3-5複素スペクトルの内挿法の原理

- (a) 複素スペクトルによる位相特性
 (b) 逆数成分の内挿による同定

$$\bar{f} = m + \frac{(u, z_{m+1})}{(u, z_{m+1}) - (u, z_m)}$$

ここで、 (\cdot, \cdot) は内積を表す。また、振幅Aは次式で推定される。

$$\bar{A} = \frac{\pi(f-m) \cdot (u, z_m)}{\sin \pi f}$$

この手法を合成は系にて起用したところ、0.1%以下の周波数誤差で、基本周波数を抽出することが可能であった。半音間の周波数の差が約6%であることを考慮すれば、この方法は採譜という目的には適していることがわかる。

3-2-2 音階の単位系

音楽で使われる音階はオクターブ（周波数が2倍までの範囲）をどのように分割するかによって、純正音階、ピタゴラス音階、平均率音階などに分類されるが、通常は最も数学的に取り扱いやすい等分平均率音階が用いられる。これは1オクターブを対数的に

1/2等分するもので、隣り合う音（半音）の間の周波数比が 1.06 （2の12乗根）となっている。言い換えると、半音間には約6%の周波数差しかないということになる。このような音高の差を数学的に表現する単位としては、2つの音の周波数比 f_1, f_2 を用いて、

$$\text{音高差 [セント]} = 1200 \log_2(f_1/f_2)$$

が、用いられることが多い。本来、音階や音程は音の高さの相対的測度であるが、楽器の調節やオーケストラのピッチ合わせなどでは標準となる周波数が必要である。現在では、A4を440Hzにした国際標準が用いられている。本論文では、C0が0セント（ピアノの中央の「ド」（C4）が3600セント）となるように絶対音階を定めている。

3-3 多重音の採譜システム

この節では、単一楽器によるポリフォニーを採譜するシステムについて述べる [Katayose 1990 a]。ここでは、人間の専門家にみられる知識処理に考察を加えることにより、従来の採譜システムに見られる制限をかなり軽減している。このシステムは、音楽感性情報処理システムの中核として、データベースの構築、音楽理解の情報知覚部として実働している。実験としては、ピアノ、ギター、三味線を対象楽器として採譜処理を実現し、本システムの有効性と限界を検討している。

3-3-1 システムの概要

音響信号は、適当な遮断周波数をもった4次のバタワース・ローパスフィルタを通して後、12ビットのA/D変換器によってサンプリング周波数10kHz、または、20kHzで計算機に格納される。サンプリング周波数が10kHzの時の入力信号の最大周波数は、サンプリング定理により5kHz以下に規定されてしまい、これは聴覚特性からいうと音色の変化がまぬがれないが、単一楽器の採譜という目的からは、十分な条件である。取り込まれた音響信号に対し、約100msの時間窓を50msずつ移動させながら、先述の複素スペクトルの内挿法を用いて、ピーク周波数ならびにその振幅の計算を行う。このようにして音のピッチとそのときの強さの時間方向に描いていく時間周波数マップ（図3-6）を得る。採譜処理は、時間周波数マップを解釈するという形で行われ、音高、音の開始時間、終了時間を記述するノートデータが抽出される。また、ここでは音の大きさ、正規化時刻に対する演奏タイミングのずれに関する計測も可能である。処理構成を図3-7に示す。

柔軟な処理を目標とする知識処理部はパターン駆動型のOPS83言語で記述し、ピーク周波数の抽出、音の立ち上がり時刻の検出など信号処理として必要な手続きの決まった処理はC言語で記述している。図3-7において知識処理部は制御モジュール、処理モジュール、音楽解析モジュールから成り立っているが、このほか、後述の高調波テーブルの作成、パラメータ設定のために楽器音解析モジュールを用意している。

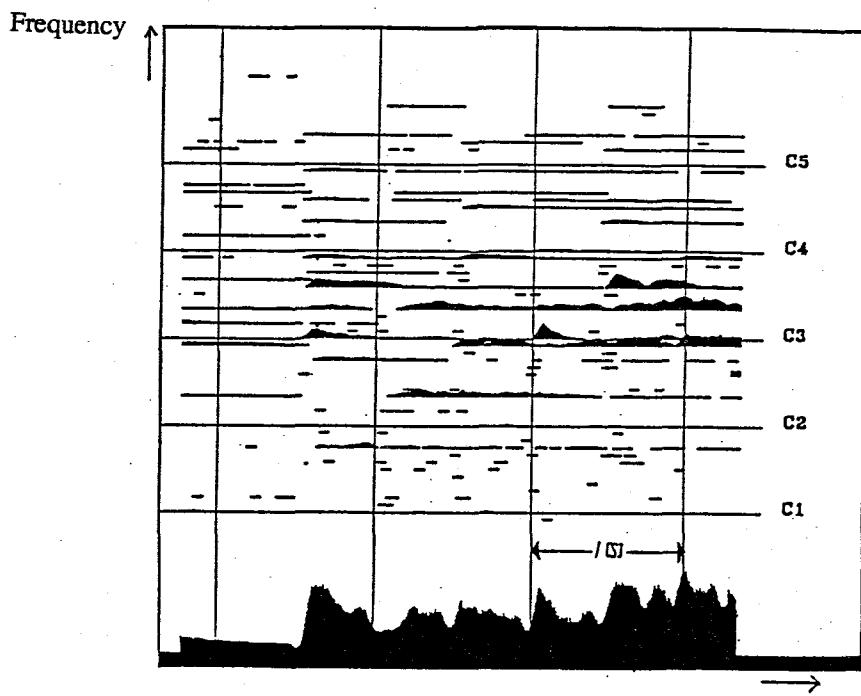


図3-6 時間周波数マップ

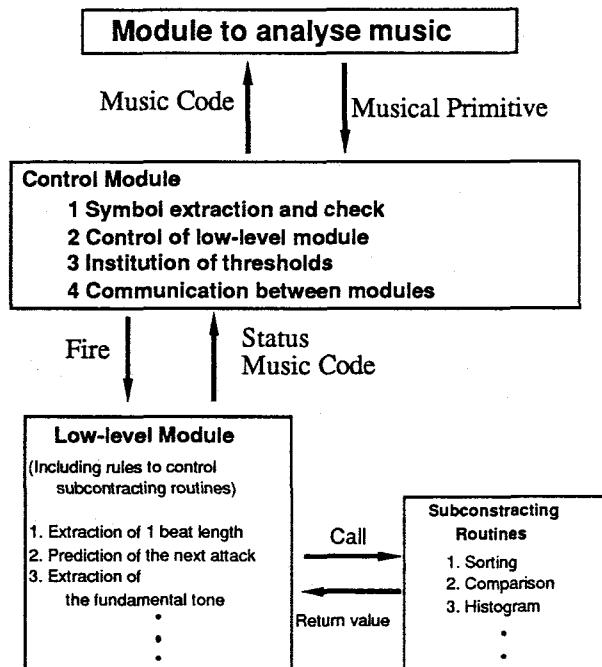


図3-7 処理の構成

3-3-2 各モジュールの説明

1) 制御モジュール

制御モジュールは本システムの中心となるモジュールであり、処理の流れを作り、処理モジュールの手続きを発火させるルール、および、音楽解析モジュールとの通信に関するルールが記述されている。本モジュールは音楽解析結果を処理に反映させるという役割を担っている。例えば、楽節の終わりには“間”が持たれるが、音楽解析モジュールの拍子解析結果をもとに4小節目、8小節目の終わりには拍の長さが延ばされる可能性があるという情報を処理モジュールに伝えている。

2) 処理モジュール

本モジュールは制御モジュールの下位に相当するモジュールであり、実際に採譜処理に必要な手続きを行なう。直接に時間周波数マップをアクセスする処理群はC言語で書かれており、これらをOPS83で制御する形で1つのモジュールが実行される。各ルーチンで必要なパラメータはルールベースで調整される。以下、中心となる処理について述べる。

3) 音楽解析モジュール

音楽解析モジュールではメロディの解析、拍子の解析、コード進行の解析を行っている。これらは音楽感性情報システムにおける音楽解析処理部である。これについては第5章で述べる。

3-3-3 ビート・トラッキング

本処理におけるビート・トラッキング（拍の量子化）は、基準となる拍の長さをまず検出し、次拍の来るべき位置を予想するという方法を用いている。具体的には、一拍の長さ、ビートの強さ、小節などの音楽的制約という情報をつかって、ビート・トラッキングは実現されている。拍の量子化の方法の多くは、曲全体的な平均的な拍の長さに注目するのに対し、本手法では、次拍の来るべき位置をトップダウン的に予想するという点で、人間の聴取過程に近いものとなっている。また、この方法は、他の研究のように理想的に音の立ち上がりが検出されていることを前提としておらず、多くのノイズを含

む音の立ち上がり候補の中から真の音の立ち上がりを検出する方法として有効である。基本波、高長波の判別、持続時間の推定などの音名の同定は、検出された拍のユニットに対して行なわれる。

1) 基準となる拍の推定

まず、音楽開始10秒間のパワーを微分して音の立ち上がり候補を抽出し、その候補間隔のヒストグラムを作成する。ヒストグラムの作成する様子を図3-8に示す。

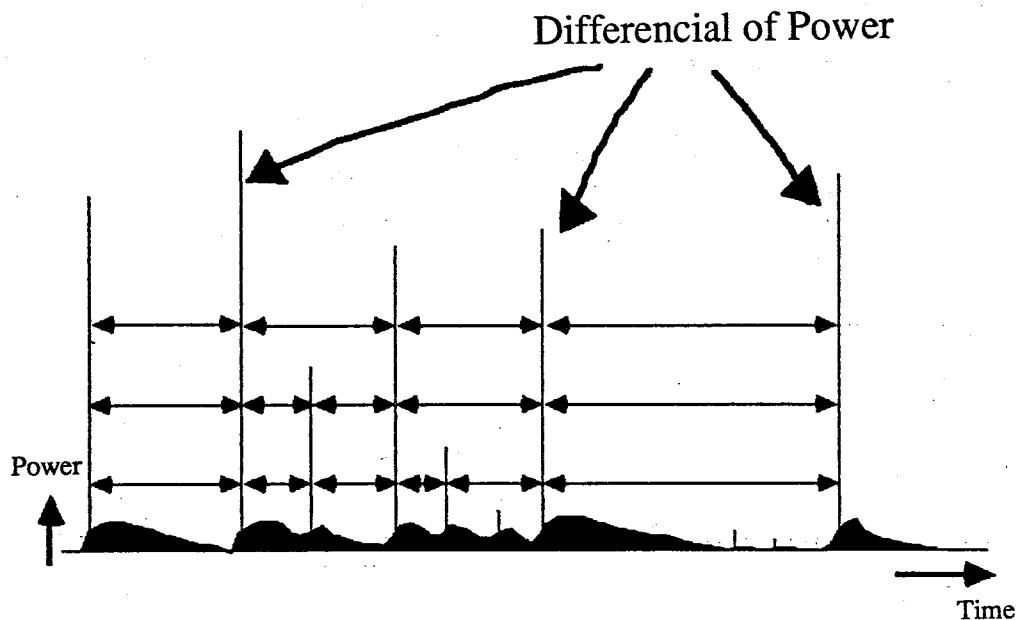


図3-8 基準拍の推定

人間が一拍として自然に感じる長さ（0.25～0.8秒）にヒストグラムのピークがあればその時間を基準拍の長さとして採用する。もし、それより短い時間にピークがあれば、その整数倍でかつ人間が自然と感じる範囲でのピーク値を検出し、その時間を基準拍の長さとする。そして、その整数倍分を分割フラグとする。分割フラグはセグメンテーションされた拍を、何分割して立ち上がりを認識するかに用いられる。これは、基準拍が短すぎるとテンポの揺らぎのため拍の認識が困難となるということを考慮した作業である。

2) 次拍の予想

拍のセグメンテーションは基準拍の長さを用いた次拍の重み付け探索によって行なう。

この様子を図3-9に示す。音の立ち上がり候補は、パワーを微分して抽出したものである。ペナルティには次拍の予想位置と音の立ち上がり候補の位置のずれを用いている。もし、明確な立ち上がり（パワーの微分値がthresholdより高いもの）が検出されないときには、その次の拍に対して同様の処理を行う。この処理を何回か繰り返して明確な立ち上がりが発見されないときはthresholdを順次下げて、探索を繰り返す。

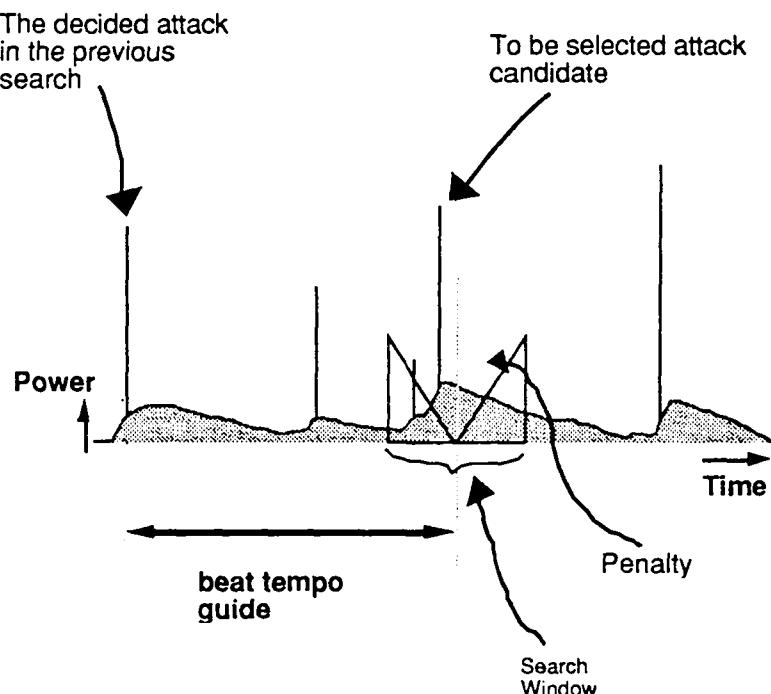


図3-9 次拍の追跡

固定した基準拍では、狂詩曲のように速度の変化が大きい楽曲に対処できないため、移動平均を用いて基準拍の長さを更新している。以下の処理は各ユニット（検出された拍）について行われる。

3) 基準拍より短い音の認識

明確な音の立ち上がりが基準拍の1/2, 1/3, 2/3の位置に検出されたとき、および分割フラグが立っているときに各ユニットを強制的に分割して、以降の認識を行なう。なお、本システムでは、5連符や2拍3連符などには対応していない。これらの認識を優先させると、ノイジーな対象に対しては余分な音の立ち上がりを抽出してしまうことが多い。

今回は、対象とする楽曲が5連符などを含む複雑なものではないことと、ある程度ノイジーなデータに対応したいということから、音楽的制約として、5連符や2拍3連符を認識の対象から外している。

3-3-4 音名の同定

1) 基本波の抽出

まず、対象とする楽器音の基本波に対する高調波のパワーの比のテーブルを用意しておく。基本波か高調波かを判別するしきい値は注目する音の周波数を高調波にもつ時間周波数マップ上の音素のパワーにテーブル中の値をかけたものの総和で与える。基本波の抽出はユニット内の音の大きさ順に楽器の種類に応じたノイズレベルまで行われる。低音部においては、基本周波数より、第2、第3高調波のパワー成分の大きい楽器もある。低音部に対しては、ノイズレベルを低く設定して探索を行なっている。

2) 立ち上がりの認識

本システムでは音の立ち上がり候補前後のパワーの比較、音の減衰比の比較により立ち上がり音、持続音の判別を行う。閾値は楽器に応じてシステムが選択する。

3-3-5 ノートデータから楽譜への変換

楽譜の生成は図3-10に示すような過程で作成される。

単なる音の羅列であるノートデータ（note形式）はルールベースで音楽的解析を行うことにより、楽譜中の各图形に対応する楽譜シンボル（scr形式）に変換される。楽譜シンボルは位置決めを行った後アポロ・ドメインのグラフィック・パッケージ2D・GMRを用いた图形に変換され、楽譜として表示される。

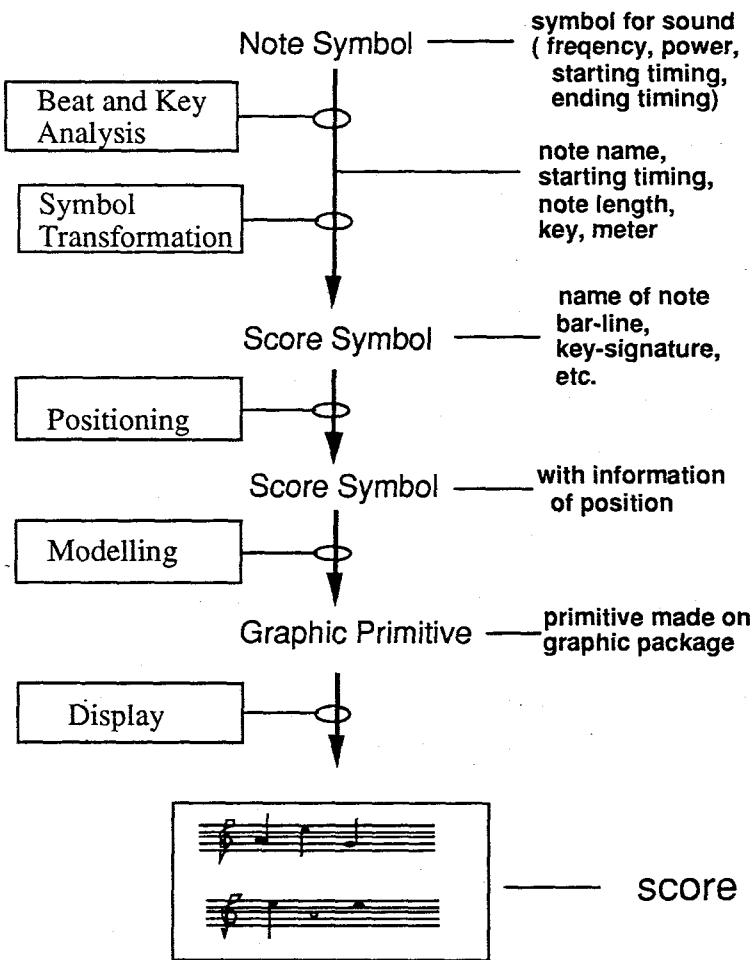


図3-10 楽譜の生成

3-3-5-1 ノートデータの解析

まず、前処理としてノートデータに対して、拍子、主調の解析を行なう。拍子および弱起（曲が弱拍で始まること）は各拍子の1拍目が強拍であるという知識を用いて検出される。主調は、メロディ中に使われる音の頻度、およびコード進行から解析される。これらの処理は音楽解析モジュールで行なわれる処理である。これについては第4章で述べる。

前処理のすんだノートデータはすべて、各音符に対応したシンボルに変換される。このうち、後に示すシンコペーションの表現、和音・連符の生成のルールにマッチしないものが単音としてそのまま楽譜描写時まで残される。広域にわたる楽曲を扱う時には各

音符に対する音部（ト音記号標記など）の割当が問題となるが、原則的にはC3以上の音符に対してはト音、B2以下の音符に対してはヘ音を属性として割り当てている。また、ピアノ譜を対象とするときには

- (1) 右手はト音部、左手はヘ音部にわりあてる
- (2) 同時に片手で届く範囲は1オクターブである
- (3) 同時に片手で抑えられる音は4音である
- (4) 片方の音部で4音の和音があるときににもう一方の音部に休符があつてはならない

をヒューリスティックルールとして付加している。

1) シンコペーションの表現

例えば図3-11aに示すようなデータは、リズムを明確に表現するために図3-11cの様に音符（休符）を分割する必要がある。これに対しては、「整数拍の音符が整数拍の部分で終わっていなければ整数部分で分割しタイを表すシンボルを挿入する」というようなルールを使われる。厳密には、 x/y 拍子の楽曲に於て、n分音符の長さを持ったノートについて、その終わる位置が y/n の整数倍拍に一致しなければ分割するというルールを用いている。

1 (G3 1.5)	5 (C4 1.5)
2.5 (A3 1)	6.5 (B3 1)
3.5 (G3 0.5)	7.5 (A3 0.5)
4 (A3 0.5)	8 (B3 0.5)
4.5 (B3 0.5)	8.5 (C3 0.5)

a)

図3-11 リズムの表現

2) 和音、連符のマージング

楽譜においては和音、連符などを一つのシンボルとして捉えたほうが自然である。和音については同じ時間になりはじめて、かつ同じ長さの音であればマージングを行うというルールを用いてシンボル化する。連符については、各拍子において8分以下の長さの連続音があればマージングを行うというルールを用いてシンボル化する。

3) シンタックスから補うシンボル

楽譜上に描かれたシンボルの中には小節線、五線、タイのようにノートデータのシンタックスから補われるものが多い。その代表的なものは休符である。休符シンボルは次のような条件で生成される。

- (1) ある区間において無音である
- (2) ある音が鳴り始めるときに鳴り終わる音がない

3-3-5-2 楽譜シンボルの図形化、表示

楽譜シンボルは楽譜に記述される図形に対応するものであるが、楽譜として出力するには各図形の位置決定と形状決定を行う必要がある。

1) 位置決め

楽譜シンボルは相対的な位置関係を示すものであるが、この時点で絶対位置の情報は持っていない。ここでは、音符の長さに応じた座標を決定し、楽譜シンボルの属性として付加するという作業を行っている。

2) 形状決定

以上のようにして得られた楽譜シンボルの名前、座標を図形作成ルーチンに渡すことによって楽譜が作成されて行く。形状に関する表現および臨時記号の挿入は図形作成ルーチンの処理として実現している。

3) 表示

本システムでは作成される図形はすべてのレベルで階層的に記述されており、その各レベルを表示することができる。また表示画面の平行移動、拡大、縮小、画面のハードコピーが可能である。本システムを用いて生成される譜例を図3-12に示す。



図3-1-2 表現可能な音符

4) 残響音の音楽的補正

ピアノにおけるペダルの使用時のように残響音の多い演奏においては実際の楽譜に記載される音の長さよりも長いノートデータが抽出される。このようなときには、コードの解析を行なったうえ、和声学的に合致しない音が響いているときには強制的に残響踏して判断するようにしている。

3-3-6 処理手順および実験

本システムの処理手順を以下に示す。システムを立ちあげると対象楽器の種類の入力を求めてくる。現在登録している楽器音モデルはピアノ、ギター、および三味線である。例えば、ここでピアノと入力するとピアノ音の高調波テーブルおよび、各種パラメータがセットされる。

時間周波数マップに対し、前処理として低音部の強調（音高が違う同じ大きさの音が時間周波数マップ上で同じ大きさになるような補正），各周波数の平均12音階への割当を行った後、前節までに述べてきた処理を適用して採譜を行う。前処理のうち前者は低音部の基本波成分を見逃さないようにするための処理である。

3-3-6-1 ノートデータの抽出

抽出されたノートデータの例を図3-1-3に、過程における認識率を表3-1に示す。

```

Header {
VL ON
UNDEFINED_UNIT 48
}
0 (A3 4 16)
1 (B3 2 25) (G#3 2 12)
2 (C4 6 22) (G3 2 18)
3 (D4 3 17) (F#3 2 12)
4 (A3 4 14) (E4 1 7) (D2 1 7) (D3 1 5)
5 (A2 2 13)
6 (D3 2 25)
7 (E3 2 17)
8 (D4 4 18) (A3 4 15) (D3 4 23)
12 (G#3 3 21) (D3 3 35) (C4 2 14) (E2 1 7) (E3 1 7)
13 (B2 1 21) (E2 2 7)
14 (E3 3 22) (B2 1 22)
15 (F#3 2 19) (B2 3 25)
16 (G#3 4 21) (D3 4 28)
17 (E3 2 5)
18 (B2 2 8)
19 (E3 1 3)
20 (G3 3 23) (B3 1 8) (C3 1 16) (B4 1 6) (E3 1 7)
(E4 1 3) (A1 2 6) (A2 2 6)
21 (E3 2 16) (E2 3 14) (G#4 2 1)
22 (B2 2 24) (B3 1 11)
23 (E3 1 16) (C3 1 14)
24 (G3 3 28) (E3 1 21) (B2 1 19) (C3 2 6)
25 (B2 2 38) (E3 2 36)
26 (C3 1 23) (C4 2 9)
29 (E2 1 16) (B2 1 14)
30 (E2 2 16) (B3 1 8) (B2 2 15) (F#4 1 5)
31 (E3 2 20) (C3 1 15) (C4 2 5)
32 (A3 3 27) (B2 4 47)
33 (B3 1 27) (G#3 2 22)
34 (G3 2 31) (C4 6 25) (E2 2 14)
35 (D4 3 19) (F#3 2 19)
36 (A3 4 16) (E4 1 8) (D2 1 9) (F3 1 6) (D3 2 6)
37 (A2 2 12) (F3 3 7)
38 (D3 3 31)
39 (E3 2 15)
40 (A3 4 22) (D4 2 18) (F3 1 11)
41 (F3 3 23)
42 (D4 2 5)
44 (G#3 3 21) (C4 3 16) (D3 3 28) (E3 1 7) (E2 1 6)
45 (B2 2 20) (E2 2 7)
46 (E3 1 14)
47 (F#3 2 20) (B2 3 26) (E3 3 15)
48 (G#3 4 22) (D3 4 30)
50 (B2 2 8) (E3 2 2)
52 (G3 2 24) (E3 1 26) (B3 2 11) (A1 3 12)
(C3 1 11) (A2 2 11)
53 (E3 1 27) (E2 3 16) (G#4 2 2)
54 (B2 2 26) (B3 1 10)
55 (C3 3 18) (C4 1 7)
56 (G3 4 32) (B2 3 18) (E4 1 8)
57 (B3 1 9)
58 (C3 2 17) (C4 2 7)
59 (E2 2 16) (B3 1 7) (B2 1 6)

```

図3-1-3 抽出されたノートデータ

表3-1

Error Type
E1 : Error in note separation
E2 : Under-extraction and over-extraction of attacks
E3 : Missing of small notes

Instrument	Music	Recognition Rate	Error Rate		
			E1	E2	E3
Piano	Gymnopedie 1	100%	0%	0%	0%
Piano	Fujikomerareta Hikari to Kage	98%	0%	2%	0%
Piano	Kaerimichi	92%	3%	4%	1%
Piano	Longing Love	82%	10%	8%	0%
Piano	Kirakira-Boshi	97%	3%	0%	0%
Guitar	Alhambra	81%	11%	1%	7%
Shamisen	Sakura	94%	2%	2%	2%

音の終了時刻に関しては音響そのものを正しいとするか楽譜を正しいとするかで評価が変わる。この段階での認識率は

(正解の打鍵音の数) - (余分に認識した打鍵音の数)

(実際の打鍵音の数)

で定義している。なお、楽器音モデルは「ジムノペディ1」を対象として調整したものである。表3-1において「Longing Love: G. Winston作曲」以外のピアノ曲に関しては認識率がほぼ90%をこえていることから通常の速さのピアノ曲に対してはほぼ認識できることがわかる。「Longing Love」における誤りのうち音の分離に関する誤認識は、低音部から高音部へ少しずつずらして奏でる和音の高音部を高調波と認識してしまうものであった。この楽曲は全体的に低音部を強く演奏する傾向があり、高調波テーブルをこの楽曲に合わせたものにすれば認識率は数%改善されるであろう。その他の誤りはこの一つ一つ音の長さが短いためにに起こる誤りである。このような楽曲においては切り出されるユニットが短くなり、その結果、立ち上がり比や、減衰比を計算する際にノイズの影響を受けやすい。音符の細かいいい楽曲やテンポの速い楽曲を対象とするときには時間周波数マップの時間方向における分解能をあげる必要がある。「アルハンブラ宮殿の想い出」(ギター)に対する誤認識はこの楽曲がトレモロ奏法を用いているためにおこる音質の変化と弦楽器特有の音の立ち上がり時に生じる摩擦音に起因するものである。

制御モジュールの説明部分で楽節のおわりの間について触れたが、ここでは“間”的割合を基準拍の10%に設定して実験を行った。これは「封じ込められた光と影：村松健作曲」に対して経験的に設定した値であるが、この値は（ここで実験を行った）他の楽曲に対しても有効であることが確かめられた。また、このほかに、基準拍より長い音符はより長めに演奏されるということが実験過程より共通事項として確かめられた。今回は次拍の重み付け予想という処理に吸収されたため、あえてルール化を行わなかったが、今後、より情感のこもった楽曲を対象とするときには参考になるものであろう。

3-3-6-2 楽譜生成結果

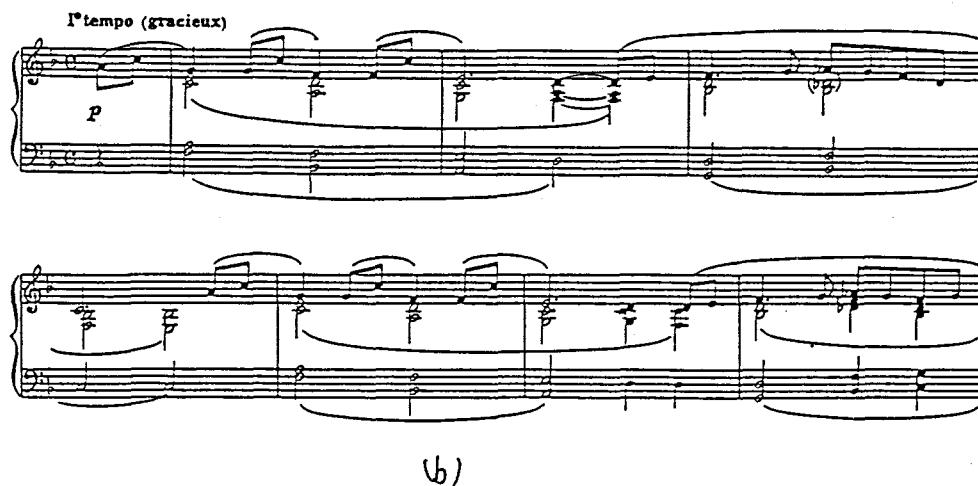
「封じ込められた光と影(図3-13)」の楽譜出力例を図3-14に、「君について：村松健作曲」の自動採譜結果と印刷楽譜との比較を図3-15に示す。ここに示した楽譜例は、音の持続時間の補正という点で不十分な所もあるが、多重音を扱って楽譜生成まで至ったシステムとしては満足の行くものである。



図3-14 封じ込められた光と影（図3-13）」の楽譜出力例



(a)



(b)

図3-15 「君について：村松 健 作曲」の自動採譜結果と市販楽譜との比較

a) システムの出力, b) 市販楽譜

3-3-7 多重音の採譜システムのまとめ

採譜とは信号処理に基づいた処理であるが、人間が採譜を行なうときに様々なフェーズで音楽的制約（知識）を用いている。ここでは、そのような音楽的制約の利用を考慮した採譜システムについて述べてきた。本システムで実現できた事項および課題を以下にまとめる。

1) さまざまな楽器（多重音）への対応

音のモデルを登録するだけでいろんな楽器に対応できるようになった。現段階では、明確なアタックを持つ楽器を対象としたが、これは微分処理で拍の検出を行っているからである。音の立ち上がりがはっきりとしない楽器の拍を発見するには、音のエンベロープモデルと時間周波数マップ上でのエンベロープを畳み込む方法などが考えられる。

基本波の抽出には近似的な方法を用いたが、ここで対象として楽音に対してはほぼ満足の行くものであった。これは実際の音響に対し、10KHz以下でローパスフィルタリングをおこなっていることと大きく関連していると考えられる。もし、本手法を用いて基本波の抽出により結果が得られない対象があれば、遮断周波数をあげるとともに、もっと精密に連立方程式を解く手法を実現してやる必要がある。これには、ニューラルネットワークを用いる手法が、パラメータの生成という点においても有望である。

2) 音楽的柔軟性

拍、音楽的段落の解析を行うことによりテンポの揺らぎ、“間”に柔軟に対応できるようになった。また、“典型的なワルツでは3拍子目の拍が短くなる”という規則があれば、本システムではそれを制御モジュールに付加するだけで簡単に採譜処理に反映することができる。

3) 楽譜の生成

レイアウトに関しては改善するべき点もあるが、構造解析を行うことにより、かなり見やすい楽譜を作成することができた。単なる楽譜生成のツールとしての機能についてはほぼ満足の行く結果が得られたが、採譜結果を出力するための知的さには検討の余地を残している。特に、持続音の音楽的補正に関しては、残響を認識するためのルールが一つ記述されているに過ぎず、ルールの拡充が望まれる。この問題を根本的に解決するには音楽の本質のルール化が必要となる。

3-4 複数音源の分離

前節では、採譜における知的処理に鑑みながらも、実用性を目的を中心に据えた自動採譜システムについて述べてきた。そこでは認識率を確保するため、対象を単一楽器による音楽に制限している。本節では、音楽の初期知覚としての認知科学的な考察に基づいた、より、制限の少ない異種楽器による複音の採譜システムについて述べてゆく。

人間は様々な情報が重積された一次元の音響信号から音源の種類、その大きさ、方向という情報を分離し、認識を行なっている。この能力は情報収集、特に危険信号の察知という面において非常に重要な役割を果たしてきた。この工学的実現は音楽分野における利用だけではなく、ロボティクス、セキュリティという分野で期待されるものである。今までの音源分離の研究は、複数マイクを用いて音源位置を同定するという建築音響的な手法で実現してきた。人間が二つの耳で三次元的情報を得ることができる理由も（通常、三次元的な情報を得るには最低3本のマイクが必要である）、独特の耳の形状という観点から説明される。建築音響学的なアプローチは、音源の分離が人間が空間的な情報により音源を識別するということに基づいたものであるが、興味深いことに、人間は空間的な情報が得られない条件においても音の分限の分離を行なうことができる。例えば、一つのスピーカーから聞こえてくる音楽に対し、「ピアノとバイオリンが鳴っている」ということがわかったり、電話の向こうから、相手の声だけでなく、どこからかけているかということがわかったりするのは、位置によらない音源分離の能力を示すものである。以下では、位置情報によらない音源分離について考えてみることにする。

3-4-1 アコースティック・アルゴリズム・コンパイラ

音源分離のためのモデルを、ここでは、アコースティック・アルゴリズム・コンパイラ(AAC)と名付けている。アルゴリズム・コンパイラとは、認識・照合手順を自動的に生成する機構を意味し、画像処理の分野では既に具体的なシステムも紹介されている[Ikeuchi 1989]。本研究における音源分離モデルも照合手順の自動的生成という処理が実現されているため、アルゴリズム・コンパイラという言葉を用いることにした。AACの最も大きな特徴は対象となる音のモデルを入力するだけで、具体的な処理が自動的に生成されるという点である。

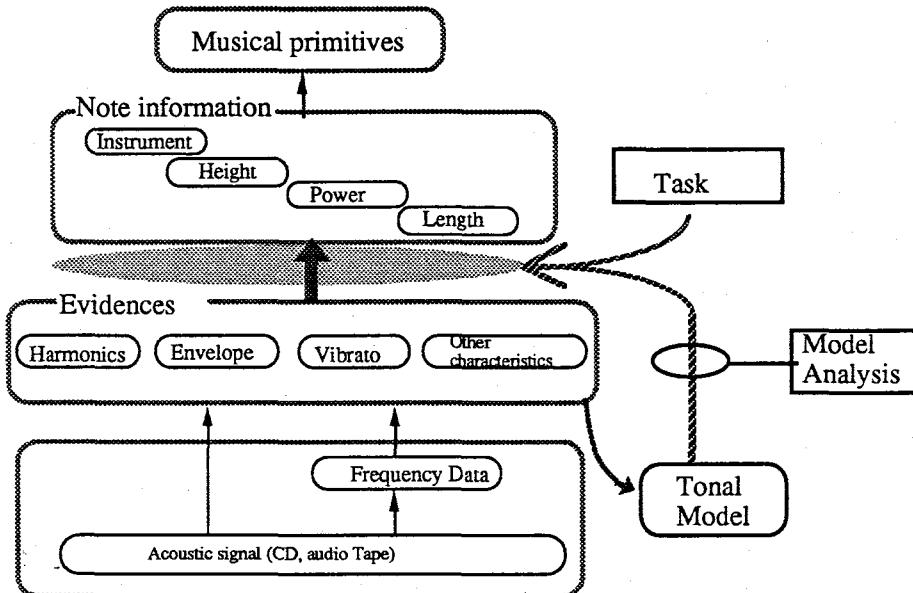


図3-16 アコースティック・アルゴリズム・コンパイラの概念図

例えば、ピアノとクラリネットによる楽曲の音源分離においては、クラリネットの周波数構造において、第2倍音が小さいという情報が生かされる必要がある。また、ピアノとバイオリンの音源分離においては、バイオリンのビブラート、すなわち、周波数の構造的なゆれが音源分離における重要な根拠情報として利用されるべきである。従来型のパターン認識システムにおいては対象に依存してプログラムが組まれてきたため、ピアノとバイオリンの分離、クラリネットとピアノの分離というような判別の根拠がことなる対象については、一つのシステムで扱うことができなかった。一方、AACに於ては周波数構造、エンベロープ、ビブラート、トレモロなどの音源分離根拠情報とモデルとのマッチングと対象となる音源のモデルの差から音源分離根拠情報に対する信頼度をシステム自身が計算するために、対象音の根拠情報に何らかの差があれば、その差に基づいた最適な処理形態が実現される。例えば、バイオリン音とピアノ音を判断する際にはビブラート成分が有力な音源分離の根拠となることが、信頼度という形で判断処理部に送られ、ビブラート成分による音源分離に対する重み付けが行なわれる。

もう一つのAACの大きな特徴はタスクモデルを利用した認識法である。タスクモデルの利用の利点はアクティブな信号処理を行なうことができるという点と認識結果の確かしさをシステム自身が知っているということである。人間は非常に調和した和声的なコーラスやオーケストラの演奏において完全に一つ一つの音の分離ができなくても、響

き（コード）としては知覚できることがある。この過程は一見何でもないことであるが、一つ一つの音が分離できないということを本人自身が判断できるのは非常に重要なことがある。パターン認識システムにおいて認識率という言葉がよく使われるが、ここでいう認識率自身は人間が測定するものである。もしシステム側が自分が行なっているタスク自身に信頼性というファクタを加味してやれば、例えばある楽曲においてベース音を追跡せよというタスクにおいて、一つ一つの音の分離に対する信頼性が低いときには、得られるコードからベース音を推測するというようなことが自立的に判断できるであろう。

3-4-2 実験

前節ではAACの概念を述べたが、現在、そのすべてがインプリメントされているわけではない。多音源の分離として考えられる対象としては、多重音の多楽器、単音の多楽器、多重音の单一楽器、楽器の推定（ある演奏においてどの楽器が使用されているかの判別）などがある。基本的には、楽器の推定と多声の単楽器（ピアノなど）と単音楽器の組み合わせに対して実験を行なえば、音源分離の基礎実験と位置づけることができる。ここでは、この二つの処理を実現できるようなAACのプロトタイプを作つて実験を行なった。

対象のモデルにはハーモニック・エンベロープ（楽器音の各ハーモニック成分におけるエンベロープ）、ビブラート、トレモロなどの演奏法の付加か否か、発音数の上限を与えることにした。現在、音のモデルとしてはピアノ、バイオリン、クラリネット、オーボエを登録している。楽器によっては、ハーモニック・エンベロープと基音のエンベロープが異なることもあるが、プロトタイプとして、エンベロープのモデルとしては、基音のエンベロープで代表させている。倍音構造については、各音高におけるハーモニック・エンベロープのパワーの周波数構成比でモデルを作成している。クラリネットとピアノのエンベロープモデル、倍音構造モデルの一部を図3-17に示す。

エンベロープモデル、倍音構造モデルとのマッチングを行なうときには音響データからハーモニックエンベロープを切りだす必要がある。時間周波数マップ上で図3-18に示すように、隣接パワー成分を切りだすことによりハーモニック・エンベロープのラベリングを行なう。ビブラート成分については、周波数データを時間方向に微分することによって抽出を行なっている。

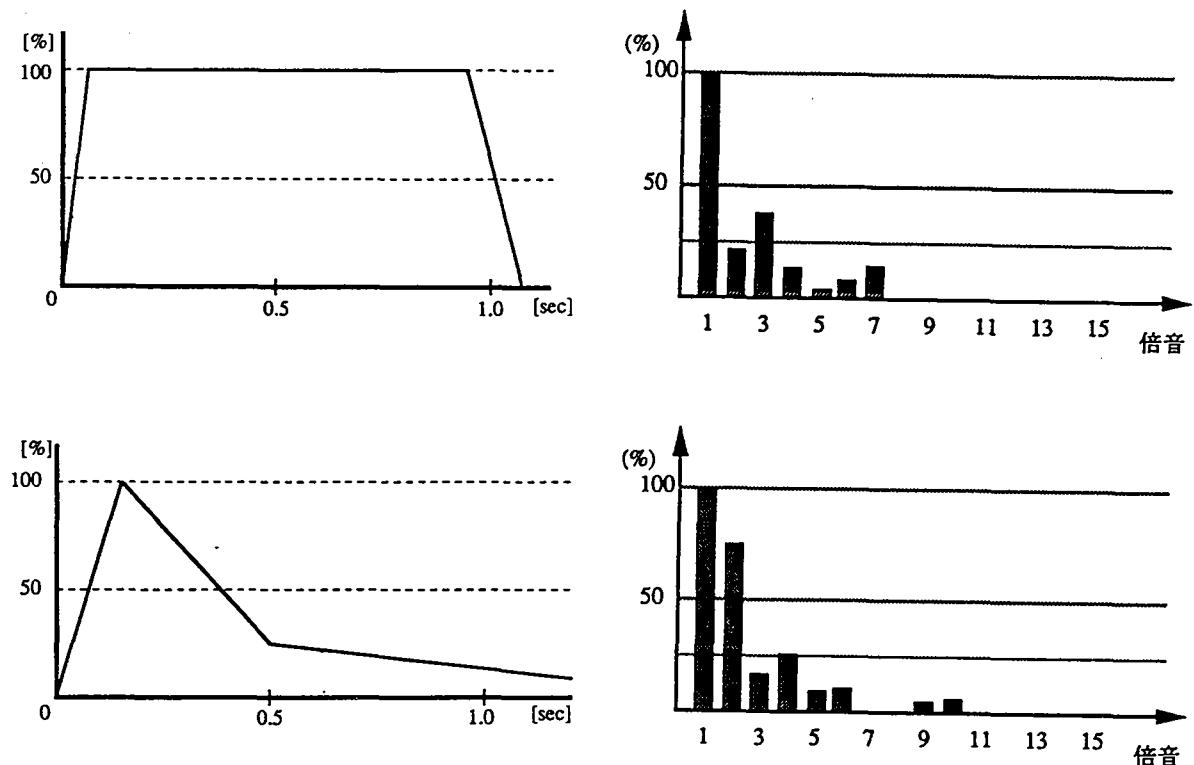


図3-17 クラリネット、ピアノ音のモデル

a) クラリネットのエンベロープモデル、b) ピアノのエンベロープモデル、c) クラリネットの倍音
モデル(C3)、d) ピアノの倍音モデル(C3)

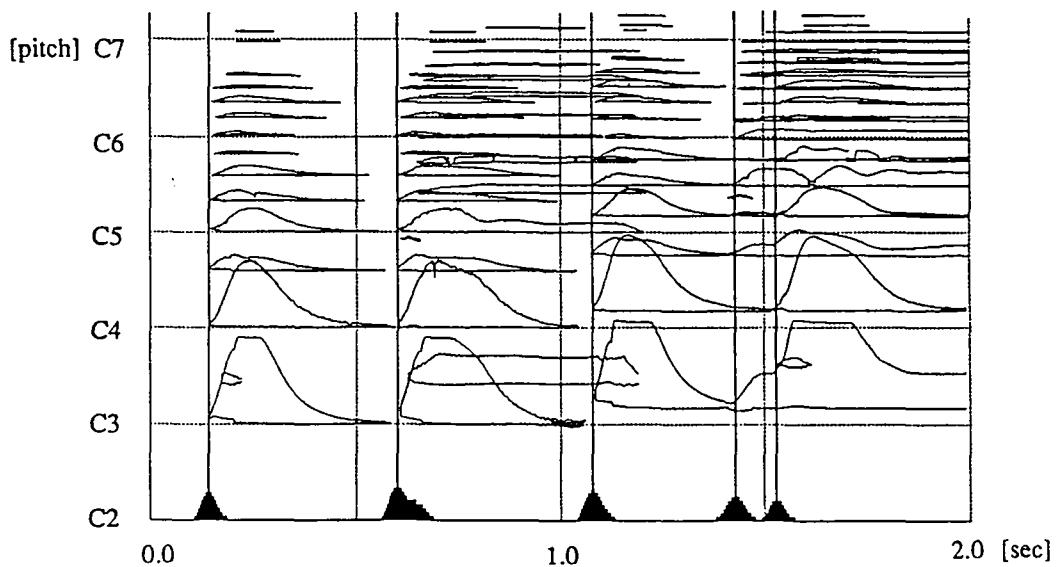


図3-18 切り出されたハーモニックエンベロープ

3-4-2-1 楽器の推定

ここでは、オーボエ音、ピアノ音、バイオリン音、クラリネット音の中からどの楽器であるかの判別を行なう。オーボエ音、ピアノ音、バイオリン音、クラリネット音の時間周波数マップを図3-19に示す。エンベロープ成分、倍音成分、ビブラート特性から見た、各音の楽器推定の可能性を表3-2に示す。これより、各楽器が判別されていることがわかる。

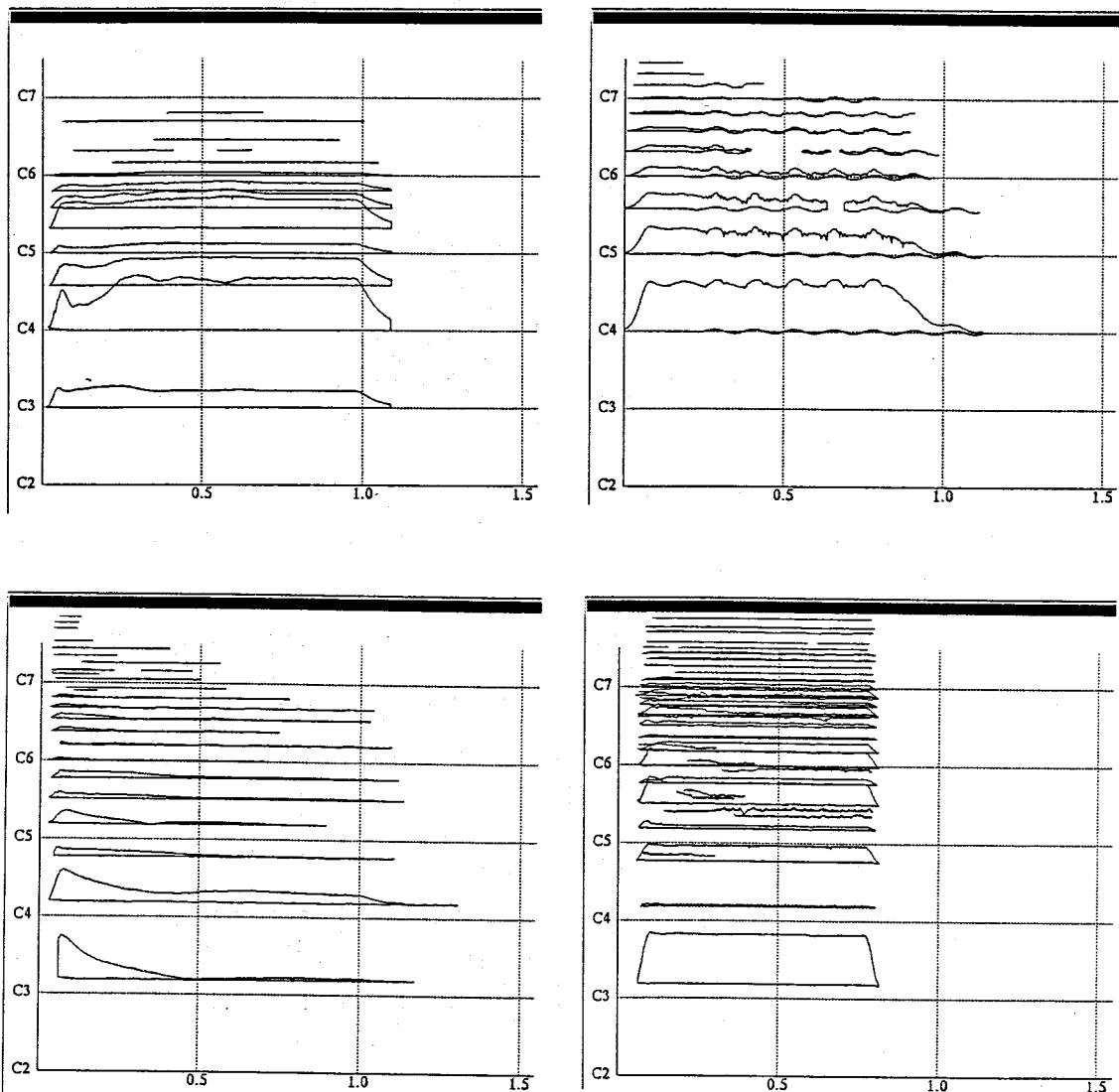


図3-19 時間周波数マップ a) オーボエ音, b) ピアノ音, c) バイオリン音,
d) クラリネット音

表3-2

Source	Envelope gravity	Harmonic Structure				Vibrato	Certainty to Source
		oboe	violin	piano	clarinet		
a) oboe	52%	98%	0%	0%	0%	0%	98%
b) piano	23%	0%	8.3%	88.2%	0%	0%	96%
c) violin	39%	0%	98%	83%	0%	100%	88.2%
d) clarinet	49%	0%	0%	0%	100%	0%	100%

3-4-2-2 ピアノとクラリネット音の分離

図3-20に実験の対象となる音響の時間周波数マップを示す。このデータに対し、エンベロープの特性から、クラリネット系の音としての確信度が高い周波数成分を図3-21a)にピアノ系の音としての確信度が高いものを図3-21b)に示す。同様に、倍音声分という観点から、クラリネット系の音と判断された基本波成分を図3-22a)に、ピアノ系の音と判断された基本波成分を図3-22b)に示す。両者ともビブラートという演奏法は含まないので、ここでは、エンベロープと倍音構造から最終的な結果を求める。この結果を図3-23に示す。この結果は、エンベロープ、倍音成分だけでは、うまく判別は行なえないが、両者を併用することにより、認識が可能になったという例を示すものである。

3-4-3 検討

本節では、自動採譜システムの構築ということから一步進んで、音源の分離モデルを立ててその基礎的な実験を行なった。現段階では、モデルの一部を実現、テストしたに過ぎないがその可能性は十分に示せたと考えている。また、周波数解析、エンベロープ、倍音構造、ビブラート成分の抽出などの処理は逐次処理として行なっているが、機構的に並列処理を意識した機構になっており、将来的にはリアルタイム処理も可能である。

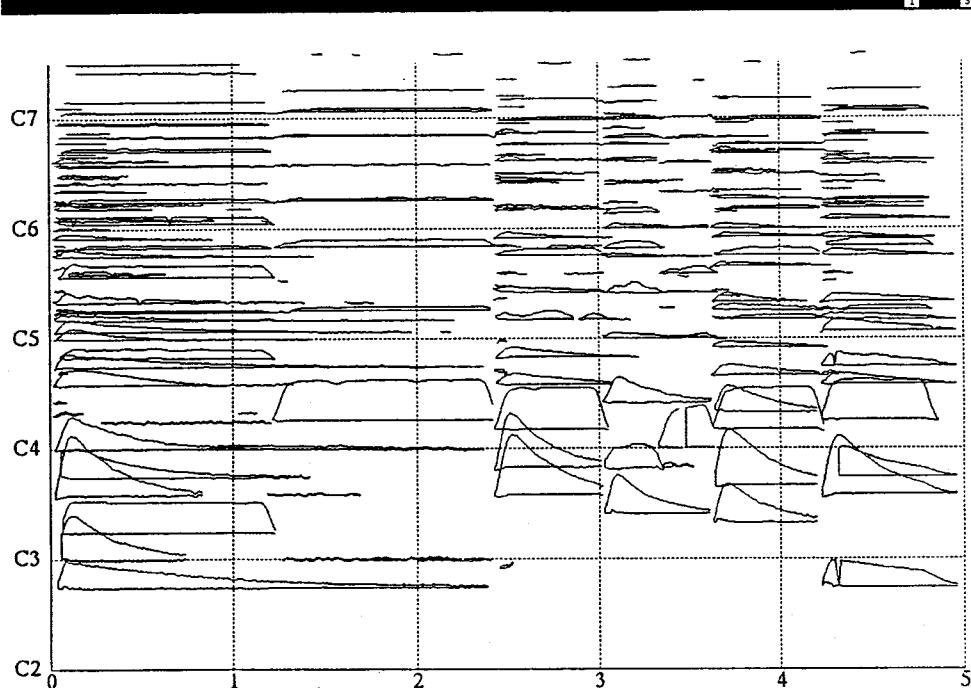


図3-20 認識対象（ピアノ、クラリネット）の時間周波数マップ

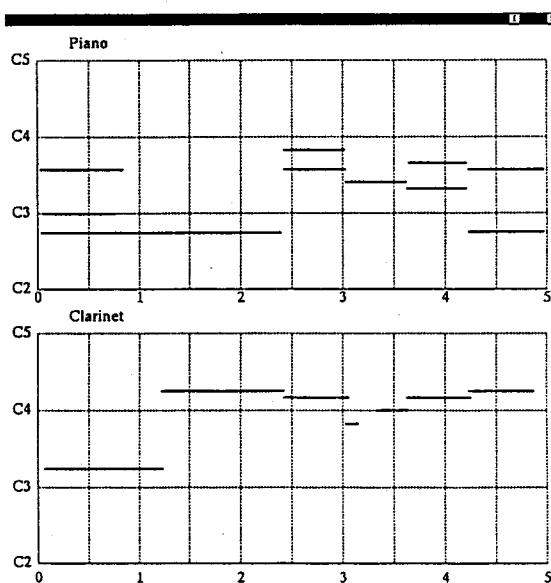


図3-21 エンベロープによる判別

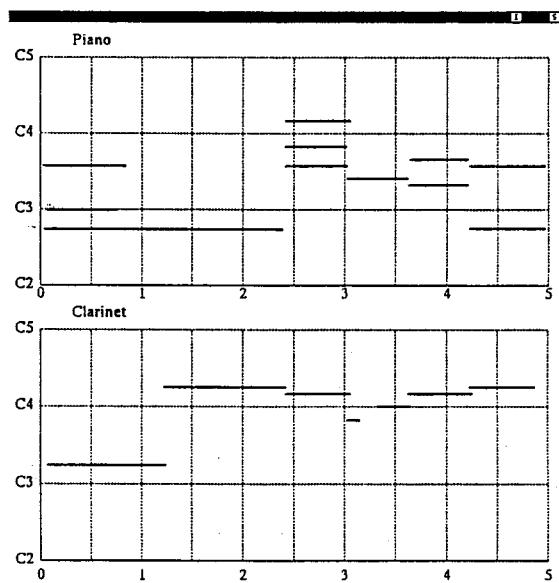


図3-22 倍音成分による判別

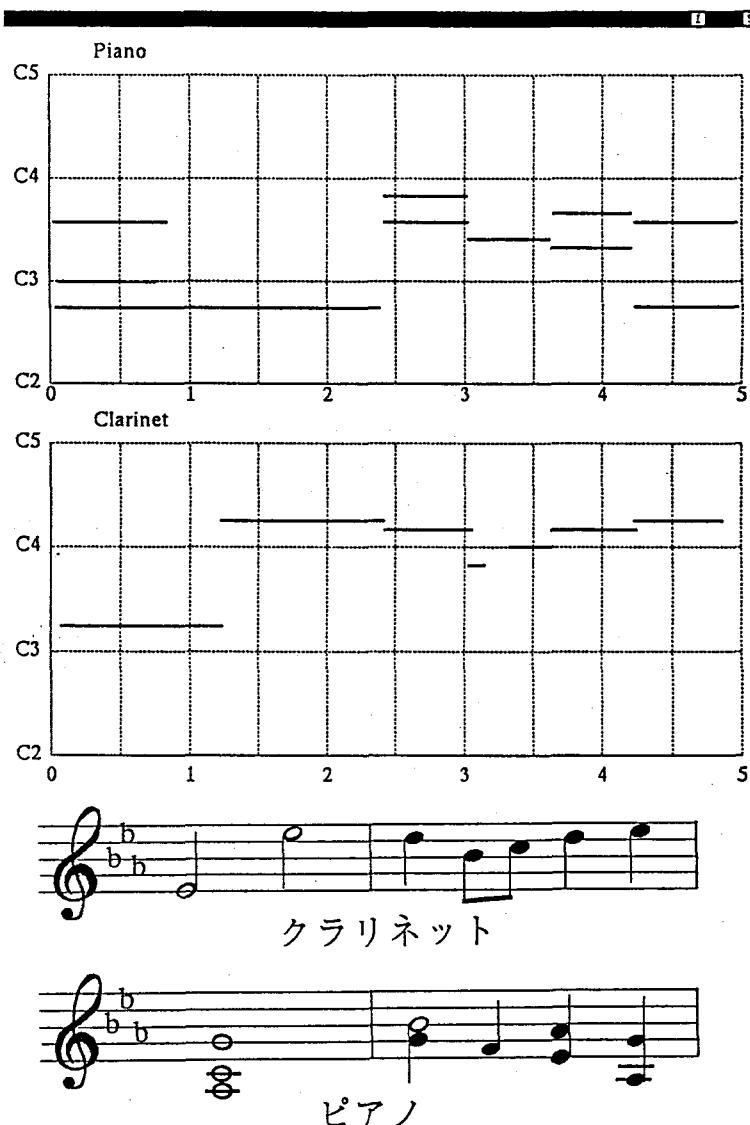


図3-23 判別結果 a) 結果, b) 対象データ

3-5 まとめ

本章では、自動採譜、音源分離について述べた。
自動採譜における諸問題を以下にまとめる。

1. 高分解能のピーク周波数の抽出
2. 基本波の抽出
3. 音源の分離
4. ビートトラッキング
5. 楽譜化

具体的なシステム例としては、実用を目指した单一楽器多重音の自動採譜処理システムと音源分離モデルAACについて述べた。それぞれの特徴を以下にまとめる。

1) ポリフォニーの採譜システム

ピアノ、ギターなどショートアタックの楽器を対象に音響信号から音列データ、楽譜を生成する多重音採譜システムである。拍、小節など音楽的制約を利用することにより音価の割当に対する精度が向上している。また、音のモデルを登録することにより、様々な楽器に適応することができる。

2) AAC (アクースティック・アルゴリズム・コンパイラ)

音のモデル、タスクモデルを利用した音源分離システムである。音のモデル、タスクを与えるだけで、自動的に認識処理が与えられるという特徴と、認識結果に対する信頼性をシステム自身が計算するという特徴を持っており、人間の初期聴覚をシミュレートした機構のシステムである。本論文では、基礎実験として、楽器の判別、ピアノとクラリネットの音源の分離について実験結果を示している。

第4章 音楽分析と理解

前章では、音響信号を入力として、楽音を認識する機構について述べた。この処理過程においては、人間に見られる楽音の知覚・認識機構をかなり実現したものとなっているが、それ自体は音響信号から楽音を抽出することが主目的であり、まだ音楽を認識・理解したという段階まで至るものではない。この章では、楽音知覚レベルより高次の音楽の理解機構について扱う。まず、拍子、調性、コード進行、メロディなど音楽を構成する要素の認識について述べる。次に、そのような音楽要素を認識したうえで、どのように感性情報としてハンドリングしてゆくかについて検討を行なう。

4-1 音楽プリミティブ

第2章では、音楽の理解とは、音楽を聞いてある種の感情（以下、センチメント）に至ることだと述べた。感情とは、元来、即物的な対象として扱えるものではなく、それを形成する処理過程をモデル化を行なう過程において扱いうるものであるとされる。その処理過程のモデルは、経験的には、非常に複雑、高度の処理系であることが想像されるにもかかわらず、心理学実験的に得られる確証的なデータというと非常に断片的なものに過ぎない。感情に対するモデリングを行なうためには、結局のところ、観察に基づくにせよ、内省的な立場でモデルを立てたうえで、検証を重ねて行くという方法をとらざるを得ない。ところが、そのモデルの問題点や誤りをみつけることはできても、正当性に関しては容易に確かめられるものではない。このような状態で、コンピュータで感情を扱うことは往々にして中途半端になり、それを試みること自体誤りであるという意見もある。しかし、感性情報処理の目的は、“人間的な感情”に対する精密なモデルを立てることではなく、よりよいヒューマンインターフェースの形として“機械的な感情”を作成することにある。本章では、「どのようにしたら、より自然に人間とコミュニケーションできるシステムを作れるか」という主題をもって、音楽理解という問題を探っていくことにする。

感情という問題に先立って実現しなければならないのは、音色（楽器）、メロディ、コード進行、リズム、演奏法などの物理的な音楽要素の認識である。少なくとも、これらを認識できなければ、音楽音響信号を単なる音、あるいは音の羅列を受け取ったに過ぎない。

ぎず、音楽として理解したとはいえるものではない。本論文では、このような音楽的因素を音楽プリミティブと呼んでいる。音楽プリミティブは物理的なデータから解析されるものであるが、それと同時に、センチメントと密接に結びついたものが多く、先述のパターン感性情報に分類されるものである。例えば、トランペット音のきらびやかさや、四七抜き音階の東洋的な雰囲気など、その理由はともかくとして、かなりの音楽プリミティブが直接的な形容詞（シンボル感性情報）と結びついていることは誰もが認めることである。したがって、物理量である音響信号、あるいはノート情報から音楽プリミティブの抽出ができるれば、簡単ではあるが、「刺激→反応」というプロダクションをもって、その楽曲に対して感想文を出力することができる[Katayose 1988b]。一方、コード進行やリズムなどがわからない音楽の未経験者が音楽を楽しんでいるという事実から、音楽プリミティブの認識を音楽理解の中間過程におくのはおかしいのではないかという意見もある。確かに、感性のすべてが個々の音楽プリミティブの単純な組み合わせから得られるものではない。が、少なくとも、音楽を楽しむ人は、程度の差こそあれ、音楽プリミティブに対する認識能力をもっているはずである。ただ、対象となる音楽プリミティブの呼び名を知らなかつたり、説明ができないために、音楽プリミティブに対する認識が行なわれていないように見えるのである。このような事柄を考慮すると、音楽感性情報処理システムは、音楽経験者より、むしろ、音楽の経験の浅い人が、感覚的に音楽を扱うためのツールと位置づけることができよう。

音楽という世界を形作る制約条件は、基本的には、音の高さ、時間的間隔であるが、それぞれさまざまなレベルのものがある。音の高さに関しては、長調、短調など普段耳にする音階構造から、半音階・全音階、さらに周波数成分、すなわち、音色そのものが秩序の構成要素となるような音楽が存在している。また、時間的にも、楽節のようなグローバルな制約をはじめ、拍子・小節から、無拍子、さらに、時間そのものがパラメータであるというレベルに至るまで、現在、非常に範囲の広い世界で、音楽が制作されている。現代音楽まで含めると、扱うべき音楽の秩序は非常にあいまいなものになってしまふため、計算機で音楽理解を扱って行くときには、対象に対するある意味で厳しい制約を課す必要である。また、情報処理として扱う以上は、いくら主観的な対象を扱いたいといつても客觀性を排除してしまうことはできない。ここでは、最も親しまれているという点から、ポピュラージャンル、または、その範疇にはいる音楽を対象として、音楽プリミティブの抽出、および、センチメントの抽出を行なうこととする。音色、テンポ等の音楽プリミティブは音響から直接的に得られるものである。以下、高次の音楽プリミティブの抽出として、コード、および、メロディ、調性、リズムの認識について述

べてゆく。なお、入力については、採譜システムで得られるデータ音高、オンセットタイム、長さ、大きさからなるノート情報（図4-1）を与えるものとする。

```

Header {
UNDEFINED_UNIT 48 <---- 基準拍の長さが0.48秒
}
0 (A3 4 16)
1 (B3 2 25) (G#3 2 12)
2 (C4 6 22) (G3 2 18)
3 (D4 3 17) (F#3 2 12)

```

図4-1 ノートデータ

4-1-1 リズムパターンの認識

リズムは音のいくつかが単位となって聞こえる繰り返し過程である。音楽においては音の高低、音の長短、音の強弱、音質がリズムの決め手となる。Longuet-Higginsはメロディを構成する音の開始時刻と長さに注目した拍子の解析法を示した[Longuet-Higgins 1884]。Longuet-Higginsの示した方法では、基本的なリズムは構造全体を2分割（2連符）あるいは3分割（3連符）していく過程で生成されるという原理と、テヌート・スタカートの割り付けられる位置に関する条件が利用されている。

一方、本論文では、音の大きさが採譜処理により測定されているので、強拍を追跡するという最も単純な方法をとることができる。一般に、音楽で使われる拍子は二拍子系、3拍子系に大別され、またビートとしては、4ビート系、8ビート系などがある。もし、拍子やビートが正しいものであれば、小節、あるいは、それぞれのビートにおけるアクセントの位置は一致したものとなるはずである。具体的には、装飾音の除く最小長さを単位とする立ち上がり時のパワーの配列を作成し、ビート n ($= 2, 3, 4, 8$) で割った剰余が一致するインデックス群の配列要素の総和に n をかけ、この結果、大きいものを選び出して行くことにより強拍のポジションを決定することができる。なお、モードが0でない音列群に最大値がある時は対象が弱起であることを示している。

4-1-2 調性の認識

メロディを入力として主調を認識するという対象は、リズムの認識と同じく

Longuet-Higginsによって研究されている。Longuet-Higginsは5度圏による音の離れ（主調からの距離）と隣接音間の距離から調性を解析する精密な具体的方法を示している。処理的には部分部分で使われている音とスケールとのマッチングをとっていることにはかならない。

本論文では、単声およびメロディがわかるときには、音高のオクターブのヒストグラムを作成し、それに対し24種の音階（主音（12）＊長短調（2））とのマッチングを行う。近親調、例えば、ハ調から見てヘ調とト調で使われる音の違いは、それぞれ、一つ分のフラット、一つ分のシャープだけである。したがって、マッチングの結果、主調とその近親調で、大きな値をもつものを選んでやることで、主調を決定することができる。

上記の方法では経過音の多い対象やメロディの抽出ができない多重音に対しては誤認識を起こしてしまうことがある。多重音についてはコード進行やテンションの解析を用いて調性の解析を行なう。まず、対象の最初のコードとテンションが許されうる調を複数個候補としてあげる。続いて、順次、その次のコードから許容できない調を消去していくことにより、調を決定するという手法を用いる。この際にGからCまたはDmからGのような解決を促す進行があるときには調性の確からしさに対する重みを与えることにより、認識の効率化を計ることができる。一旦、調性が決定された後は、転調の可能性をトレースしてゆくことが可能である。また、後に、示すように数個の音列データで調を決定できるという利点がある。

4-1-3 メロディの抽出

メロディ自体の解析等は非常に積極的に行なわれてきたが[Ohura 1989]、多重音、あるいは、複数の楽器からなる音楽におけるメロディ、あるいはサブメロディを抽出することは非常に難しいとされる。このタスクに関しては、基本的には2つのアプローチが考えられる。一つは信号処理からのアプローチである。たとえば歌謡曲のように歌のパートがメロディであるというときには人間の声だけを抽出する信号処理が実現されればメロディの抽出は可能となる。二つ目は音楽理論からのアプローチである。ピアノあるいはコーラスなどの多重音のシーケンスから音楽的制約を用いて、メロディを抽出するアプローチである。これに近い研究としては、心理学の分野において、群化の研究（旋律構造の分離）が試みられている。しかし、これはどのように多声が記憶されるを実験的に示したものであり、メロディ抽出のアルゴリズムを確立したものはない。本論文ではメロディの構造上の制約として

- ・大きく聞こえる音の列である。
- ・和音でなっているときは最も高い音がメロディの構成音である。
- ・極端な跳躍進行はない（音が飛び過ぎない）。

といった簡単な信号処理と音楽ルールを組み合わせることにより抽出を行う。簡単ではあるが、このような制約だけで、多くの楽曲に対してメロディを抽出することができた。しかし、これらのルールは物理的な要素しか考慮していないため、ベースパートがメロディを担当するような場合やオブリガートが入った楽曲など対処できないという欠点がある。メロディの知覚問題は、音楽認知における大きなパラダイムの一つとなっている。

4-1-4 コード進行の認識

コード進行とはメロディに対する伴奏部の和音シーケンスである。コード進行を調べることは楽曲の雰囲気を知るうえで非常に重要な手掛かりになる。

コードは基本的に基音とドミナント音と長短を決定することができる。原理的には、それぞれのコードの展開系をモデルとしてマッチングを行なうことにより認識できると考えられる。しかし、実際にはアルペジオのようにコードを展開して演奏を行なったり、また、テンションが加えられたり、省略音などがあるため、単純なマッチングだけでは認識できるものではない。

ここでは、局所的な範囲でコードネームを決定するステージと隣接するコードネームから統合しうるものは統合するという方法でコードの割当を行なっている。

処理の流れという面から述べると伴奏部のベース音を仮定し、それをもとにどのような半音間隔で音が積み重なっているかを探索することにより、局所的なコードネームを決定していく。もし、ベース音を仮定したときにコードを決定する音があれば、その時点でコードネームを与える。コードを決定するまでの間、ベース音が、オンの状態が持続されれば問題はないが、そうでないときには簡単にコードを決定することはできない。このようなときには、前になったベース音が、心理的に、鳴っているものと仮定し、次に続く音の探索しながら、コードの決定を行なう。この時、探索される音は、前のベース音に対するコードを決定する要素として使われるとともに、新たなベース音候補としてのフラグをつけ、コードの探索要素に加えられる。もし、探索途中でコードを決定する音が見つかった時には、途中でベース候補になった音に対し、ベース音候補のフラグをはずしてゆく。省略音という問題に対処するためには、コードを認識するための優先度が付けられている。例えば、Cの音がベース音という仮定があるときEが発見された

時点では、Cメジャーというコードが仮定される。もし、その後にAの音が来れば、Amとして認識される。照合の行なわれるコードはポピュラー音楽でよく使われるもの13個（Cを例に上げると、C C(2) C(3) C9 C7 Cmaj7 Caug Cm Cm(2) Cm(3) Cdim Cm7 Cm9）を用意している。実際は12通りの調性があるため、計156通りのコードが割り付けられる可能性がある。

上記のように局所的にコードネームを与えて行く方法では、コードが過度に分割されてしまうことがある。これに対し、BdimならばBm7に読みかえてよいというような機能の代行に関するルール、連続するコードでテンションが異なるだけなら統合を行なうというルールを適用することによって、マージングを行なった結果、通常、人間が与えるようなコードの割当が行なわれる。

4-1-5 音楽プリミティブと形容詞とのリンク

現在、コード進行、テンポ、音色などの音楽プリミティブとセンチメントとのリンクを記述するルールが約30個登録されている。これは、音楽プリミティブに対するアンケートより得たものである。この数自体は少ないが、音楽が組み合わせの芸術であることを考えると、かなりのバリエーションの感想文を生成することができる。ルールの例を図4-2に示す。

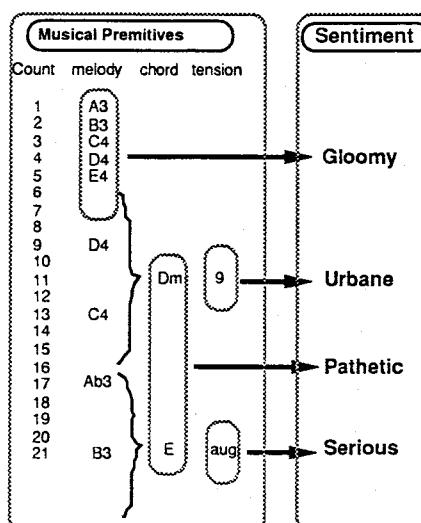


図4-2 センチメント抽出のためのリンクの例

4-1-6 実験および検討

ここでは、第4章で取り上げた村松健作曲「封じ込められた光と影」を例に解析結果を示す。この対象のノートデータを図4-3に示す。

```

Header {
VL_ON
UNDEFINED_UNIT 48
}
0 (A3 4 16)
1 (B3 2 25) (G#3 2 12)
2 (C4 6 22) (G3 2 18)
3 (D4 3 17) (F#3 2 12)
4 (A3 4 14) (E4 1 7) (D2 1 7) (D3 1 5)
5 (A2 2 13)
6 (D3 2 25)
7 (E3 2 17)
8 (D4 4 18) (A3 4 15) (D3 4 23)
12 (G#3 3 21) (D3 3 35) (C4 2 14) (E2 1 7) (E3 1 7)
13 (B2 1 21) (E2 2 7)
14 (E3 3 22) (B2 1 22)
15 (F#3 2 19) (B2 3 25)
16 (G#3 4 21) (D3 4 28)
17 (E3 2 5)
18 (B2 2 8)
19 (E3 1 3)
20 (G3 3 23) (B3 1 8) (C3 1 16) (B4 1 6) (E3 1 7)
(E4 1 3) (A1 2 6) (A2 2 6)
21 (E3 2 16) (E2 3 14) (G#4 2 1)
22 (B2 2 24) (B3 1 11)
23 (E3 1 16) (C3 1 14)
24 (G3 3 28) (E3 1 21) (B2 1 19) (C3 2 6)
25 (B2 2 38) (E3 2 36)
26 (C3 1 23) (C4 2 9)
27 (B2 2 40) (E3 1 18)
28 (E3 3 37) (A1 4 23) (A2 1 20)
29 (E2 1 16) (B2 1 14)
30 (E2 2 16) (B3 1 8) (B2 2 15) (F#4 1 5)
31 (E3 2 20) (C3 1 15) (C4 2 5)
32 (A3 3 27) (B2 4 47)
33 (B3 1 27) (G#3 2 22)
34 (G3 2 31) (C4 6 25) (E2 2 14)
35 (D4 3 19) (F#3 2 19)
36 (A3 4 16) (E4 1 8) (D2 1 9) (F3 1 6) (D3 2 6)
37 (A2 2 12) (F3 3 7)
38 (D3 3 31)
39 (E3 2 15)
40 (A3 4 22) (D4 2 18) (F3 1 11)
41 (F3 3 23)
42 (D4 2 5)
44 (G#3 3 21) (C4 3 16) (D3 3 28) (E3 1 7) (E2 1 6)
45 (B2 2 20) (E2 2 7)
46 (E3 1 14)
47 (F#3 2 20) (B2 3 26) (E3 3 15)
48 (G#3 4 22) (D3 4 30)
50 (B2 2 8) (E3 2 2)
52 (G3 2 24) (E3 1 26) (B3 2 11) (A1 3 12)
(C3 1 11) (A2 2 11)
53 (E3 1 27) (E2 3 16) (G#4 2 2)
54 (B2 2 26) (B3 1 10)
55 (C3 3 18) (C4 1 7)
56 (G3 4 32) (B2 3 18) (E4 1 8)
57 (B3 1 9)
58 (C3 2 17) (C4 2 7)
59 (E2 2 16) (B3 1 7) (B2 1 6)
60 (G3 4 29) (A2 2 19) (A1 3 17)
61 (E2 4 18)

```

図4-3 解析対象となるノートデータ

図4-4にコード進行、ビート、メロディを解析した結果を示す。図4-5には、この曲に対し、システムが生成した感想文を示す。ここでは、曲全体に散らばったセンチメントのうち、多いものから順にピックアップすることにより、曲を支配するセンチメントを調べ、文を作るテンプレートに形容詞を埋め込む方法を用いて感想文を作成している。また、コード進行のパターン、メロディのピーク部分から「サビ」（最も盛り上がる部分）を抽出し、特に、その部分に対する形容詞情報を述べるという処理を行なっている。

全体的に悲しげで、「サビ」の部分で希望的な部分が見受けられるという評価は、対象曲の題名をよく表現したものと言える。

今後の課題としては、扱いうる音楽プリミティブを増やしてゆくこと、具体的にはその認識法およびどの形容詞のリンクさせるかという問題があげられる。また、より複雑な楽曲を対象とすることを考えると音楽プリミティブ抽出のための知識の構造化、知識の獲得法に関する検討が必要となる。

count	melody	status	power	
1	A3	attacked	166	
2	B3	attacked	241	
3	C4	attacked	223	
4	D4	attacked	172	
5	E4	attacked	77	
6	E4			from 5 to 12 chord = Dm9
9	D4	attacked	177	from 13 to 20 chord = Eaug
10	D4			from 21 to 33 chord = Am7,9
11	D4			*** Key is Aminor ***
12	D4			from 34 to 34 chord = Eaug on A
13	C4	attacked	133	from 35 to 35 chord = C (2) on E
14	C4			from 36 to 36 chord = Em7,9
17	Ab3	attacked	214	from 37 to 44 chord = Dm9
18	Ab3			from 45 to 52 chord = Eaug
19	Ab3			from 53 to 65 chord = Am7,9
20	Ab3			from 66 to 66 chord = Eaug on A
21	B3	attacked	84	from 67 to 67 chord = Am
22	B3			from 68 to 68 chord = Bm on E
23	B3			from 69 to 76 chord = Dm7,9
24	B3			from 77 to 82 chord = G7,9
25	G3	attacked	288	from 83 to 84 chord = Dm (2) on F
26	G3			from 85 to 92 chord = C9
27	G3			from 93 to 97 chord = F9
33	A3	attacked	276	from 98 to 100 chord = G on F
34	B3	attacked	265	from 101 to 104 chord = A on B
35	C4	attacked	252	from 105 to 108 chord = Bm-5
36	D4	attacked	186	from 110 to 112 chord = Bsus4
				from 113 to 116 chord = Bm
				from 117 to 129 chord = Am7,9
				.
				.
				.

(a)

(b)

図4-4 抽出された音楽プリミティブ

hopeful caused by chord time = 69 to 97
pathetic caused by chord time = 133 to 153
pathetic caused by chord time = 37 to 65
pathetic caused by chord time = 5 to 33
melancholy caused by chord time = 105 to 116
urbane caused by tension time = 149 to 153
smart and serious caused by tension time = 149 to 153
rural caused by tension time = 141 to 148
smart and serious caused by tension time = 141 to 148
smart and serious caused by tension time = 133 to 140
urbane caused by tension time = 117 to 129
smart and serious caused by tension time = 117 to 129
rural caused by tension time = 105 to 116
smart and serious caused by tension time = 105 to 116

slow and leisurely caused by speed time = 0 to 0
sorrowful caused by key time = 0 to 0
hopeful in release time = 75 to 75
urbane in release time = 75 to 75
smart and serious in release time = 75 to 75

Whole atmosphere of this music is
slow and leisurely.
sorrowful.
smart and serious. Level is 18.
urbane. Level is 7.
rural. Level is 6.

There is pathetic mood on chord from 5 to 33.
There is pathetic mood on chord from 37 to 65.
There is hopeful mood on chord from 69 to 97.
There is melancholy mood on chord from 105 to 116.
There is pathetic mood on chord from 133 to 153.

On release at 75, there is hopeful, urbane, smart, serious mood.

図4-5 システムの生成した感想文

4-2 音楽聴取モデル

前節では、音楽プリミティブを抽出し、それから直接的にセンチメントを抽出する方法（ダイレクトモデル）について示した。ダイレクトモデルは、形容詞を使ったマンマシン・インターフェースの実現という当初の目的には満足の行くものであるが、そこで用いられる形容詞は、あくまでもタグとして用いられた擬似センチメント（Quasi-sentiment）である。システムがシンボル感性の時間および空間的シントックスを理解しているわけではなく、感性情報としての意味を構成しているものではない。本節では、もう少しマクロな観点から人間に即した音楽理解の形態について考察を行なう。

4-2-1 人間の音楽聴取の観察

まず、聴取過程のモデリングのためのマクロな必要条件を探るために、音楽を聞いた感想（形容詞）とその感想をもつに至った理由についてアンケートを行なってみることにした。厳密なものではないが、音色という要素を除外するために対象楽器をピアノとし、誰もが聞いたことのあるようなメロディに対して、コード進行、テンポ、フレーズの繰り返し方をパラメータにとってサンプルを生成し、音楽経験者、特に音楽教育を受けていないものの両者に対して調査を行った。この結果、

- 1) 個人差はあるが、曲によっては、類似した形容詞が用いられる。但し、好き、嫌いに関しては人によってばらつきが多い。
- 2) 同じフレーズが場所によっては違うように感じられることがある。特に伴奏であるコード進行が異なれば、同じメロディに対しても異なった形容詞が用いられることが多い。
- 3) 音楽経験により、興味のひかれ方が異なる。

ということが確認された。1) は、音楽要素に対する心情が、かなり客観的な観点から引き出されることを示すものである。反面、好き、嫌いに関しては他の形容詞と同様には扱うことができないことを表している。2) はメロディから評価形容詞を抽出するということに否定的な材料となるものである。現段階では、使われたスケールおよび、音の飛び具合に対して、評価を行うという程度が妥当であると考えている。コード進行やテンションといった音楽要素はある程度、音楽のシントックス情報を表したものであり、また、一般に使われる用例もかなり限定されるため、アプローチとして取りかかりやす

い音楽要素であると思われる。3)は、音楽経験者であればあるほど、特定の音楽要素に対して、「この部分が、面白い」とか、「誰かの真似だ」いうように評価するというものであった。

- 今回のインタビューでは特に出てこなかったが、聴取過程のモデル化においては、
- 4) 音楽を聞く時の状態によって感じ方が違う。
 - 5) 同じ曲でも聞く度によって感じ方が異なる。（同じ楽曲を何回も聞いていると飽きてくる）
 - 6) 嗜好する音楽が年齢（音楽経験）にしたがって変化するなどを考慮する必要があると考えられる。

4-2-2 音楽聴取モデル

ダイレクトモデルの問題点と、いくつかの観察から、

- 1) 感性状態を空間的・時間的に記述しうる
- 2) 音楽経験に依存しない感性情報（情）抽出機構、および、音楽経験に依存する感性情報（興）抽出機構をもっている
- 3) 音楽嗜好を判断する
- 4) 心理状態を統括し、言語で表現する

の事項に留意して聴取モデルを構築する。モデルの概観を図4-6に示す。この図において、伝播モデル・ファクタモデルは、音楽要素に対して素直に感性情報を抽出し、記述するものである。ここでは「情」を担当するものと位置づけている。それに対して、「興」を扱うものとして能動時に感性情報を抽出するのが予期モデルである。一般に、音楽経験が増すほど、音楽に対する予測・予期による楽しみ方の度合いが増えるが、これは予期モデルにおける知識の増加に対応するものと考えられる。総括モジュールは上記のモデルの内容の総合的評価と嗜好の判別を行うためのモジュールである。嗜好に関する部分の評定を特別に設けた理由は、この部分が、特に主観に近い要素であり、個人差がもっとも多いと考えられるからである。人間（生活環境、年令などによって若干のちがいもあるが）に、ほぼ共通な、一般性のある部分は伝播モデル・ファクタモデルで扱い、総括モジュールが伝播モデル・ファクタモデルの状態（クラスタ）に応じて、嗜好の判断を行なうという形態をとることにより、例えば、「緊張嗜好型」とか「陽性嗜好型」などの表現が可能となる。最後に、感性状態を出力するためのモジュールが表現モジュールである。

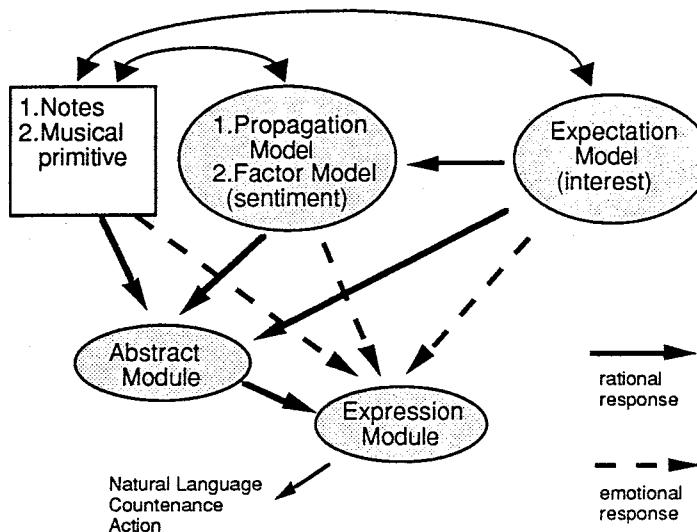


図4-6 音楽聴取モデルの概観

4-3 感性状態の記述

「楽しい」という方向に気持ちが振れると「悲しい」というセンチメントが減少するだろうし、また、「落ち着いた」という気分の時には「安心感のある」という成分も増すはずである。このようなことがらを関係付けるということは、感性情報処理にとって是非とも必要なことである。これは、従来、人工知能分野で議論されてきた「意味論」とほぼ同様の対象として扱うことができる。しかしながら、上記のように形容詞で表象されるセンチメントは、人工知能で扱われてきたような名詞・事柄を表すような通常のシンボルとは扱いが異なる。シンボルというより、スカラー量的なものである。実験心理学においては、相対的な距離をもって計測されうるものとして、実践的な研究が行なわれてきている。

形容詞的意味を扱う方法としては、(意味)ネットワークを構成する方法、意味空間を定義し、座標として概念を定義するという方法が考えられるが、本論文では、前者を1) 伝播モデル、後者を2) ファクタモデルとしてとりあげている。ともに、音楽要素も形容詞も心理モデル上に投影されるものと前提に立ち、基本的にはSD法を用いて、モデリングを行ったものである。

4-3-1 伝播モデル

使用する形容詞をそれぞれ意味単位とし、出来るだけ多く次元として用いるという立場にたったのが伝播モデルである。伝播モデルにおいては、後述の因子分析の結果、意味に相関のある形容詞が一つの意味単位としてネットワーク化される。各形容詞の類似度を対極にあるものが-1、同じ座標にあるものが1となるように正規化を行ない、この類似度をネットワークの結合度として、音楽要素に対応する形容詞の影響を伝播させてゆく。

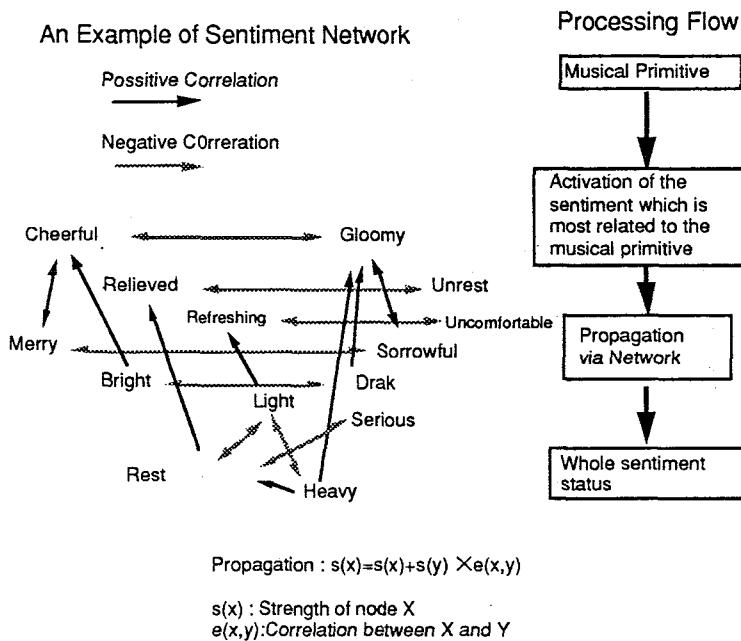


図4-7 伝播モデル

4-3-2 ファクタモデル

伝播モデルができるだけ多くの意味単位を用いるのに対し、音楽プリミティブに対する印象をできるだけ少ない次元の意味空間で定義しようという立場をとるのがファクタモデルである。形容詞対 ([Iwashita 1979]に音楽に関係するものとして上げられているもののうち「好き-嫌い」などを除いたもの、付録1参照)， 例えは、明るい<->暗い、軽やか<->重たいなど) で音楽プリミティブのイメージを評定させ、因子分析

(付録2参照)を行う。抽出された因子を感性空間の軸とし、その空間での各音楽プリミティブ、形容詞の座標を再び算出し、感性的意味のパラメータ化を行なう。各音楽プリミティブが認知される度にその時の感性座標を音楽要素の座標方向へ遷移させていくことにより、感性の時間的変化を表現する。この方法には、概念がユークリッド的に与えられるという大きな前提があるものの、小数のパラメータで対象を扱うことができ、デザイン支援などにおいて利用が進められている。

なお、評定に関しては様々な方法があるが、二対実験（二つの要素に対し、どちらが指定した形容詞において強いものかを調べてゆき、順位から評定値を割り出す方法）が推奨される。簡単には、5段階評定などという方法もある。また、嗜好に対する形容詞対を評定の対象から除いているため、評定に関してはかなりの一般性があるものと考えられる。被験者の数は多いほど安定した分析結果が得られるだろうが、個人に対して実験を行うことにより、個人の形容詞空間を求めるこどもできる。

Factor Level	a(1)		a(2)		a(3)		a(4)		a(5)	
	Primiti- ve	Loading								
positive	9	0.91	8	0.82	6	0.45	2	0.43	4	0.40
	4	0.89	1	0.60	1	0.19	5	0.28	3	0.31
	5	0.88	3	0.59	9	0.11	3	0.15	9	0.23
	7	0.68	5	0.35	4	-0.08	7	0.12	1	0.13
	3	0.53	2	0.20	5	-0.15	1	0.09	7	0.12
	8	0.08	7	-0.06	8	-0.24	6	0.01	6	0.05
	1	0.01	4	-0.16	3	-0.25	4	-0.05	10	-0.05
	2	-0.28	9	-0.20	7	-0.28	10	-0.08	2	-0.24
	10	-0.42	6	-0.26	10	-0.54	8	-0.31	5	-0.34
	6	-0.86	10	-0.60	2	-0.79	9	-0.35	8	-0.36

図4-8 因子分析の結果例

図4-8は付録3に示す音楽プリミティブを上記の形容詞対で10段階評定して得られたデータ（付録4）に対し、因子分析を行なった結果の一例である。因子分析を行なうと、因子（図のa(n)）と要素との相関の度合い（因子負荷量、図では、Loading）を求めることができる。因子は、数学的には、相関行列の固有値に過ぎない。因子の名前は分析データ（ここでは、音楽プリミティブ）との相関から、人間が決定してやる必要がある。上記の例では、因子負荷量の広がりが大きいものから順に、透明感、一般性、複雑度、洗練度、哀感が因子軸の名前として考えられる。2, 3の実験結果と感性パラメータとしてのわかりやすさという観点から、嗜好に関する部分を除外した音楽感性空間

の因子軸としては、明るさ、一般性、緊張感、洗練度の4つぐらいが適當ではないかと考えている。

4-4 予期モデル

予期モデルはある時点での展開を予想し、その結果により「興」に関する形容詞を抽出するものと定義した。これは、「予想どおり」あるいは「予想とは異なっている」というどちらかと言えば客観的な事象が「安心感」と「緊張感」および「面白さ」「退屈さ」などの主觀を想起させるという考え方の上に立ったものである。

このようなモデルにおいては、音楽のどの要素が予想できるかということが非常に重要である。予期・予測という処理を扱っていく際には、基本的には、二つのことを考えなければならないだろう。

まずは、音楽以前の、信号の変化点に覚える生理的に興味についてである。これは音楽を受容する聴覚に限った問題ではなく、視覚・嗅覚など他の感覚機能にもあてはまることがある。外界から有用な情報、特に、危険情報などは、多くの場合なにかの変化として現われるものである。変化に関する感受性は、生物が、生来、身を守るために必要不可欠な能力であり、この能力そのものが音楽聴取においても大きな要素となっていると考えられる。

もう一つは音楽知識に裏付けられた予測という問題である。音楽には芸術としての制約条件があるが、その制約条件は次に来るパターンや起こりうるパターンを規定している。ここでの制約条件とは、いわゆる禁止則だけではなく、こちらの方が美しいというような主観的な制約も含むものとする。人間は、制約条件がわかってくると、これらを知識として利用することにより、構造を解析し、次のパターンを予想しながら、能動的に聴くことができる。この予想も何種類かあり、「このパターンであれば、ありきたりである」と即断に結びつくものもあれば、「このパターンが来て欲しい」という要望や曖昧性を伴った予想などもある。筆者個人としては、予想の範囲にはいっていないのにかかわらず、心地よい進行に出会ったときに大きな喜びを感じるところである。

以上、予測という問題に対し、二つのことがらを取り上げたが、多くの研究は後者に基づいたものである。音楽学では、拍の構造、骨格音と経過音などをもって、解析、ひいては予測問題を扱っている。一方、Meyer学派に代表される認知科学サイドの研究者は、人間の精神的な作業に留意してモデルを作るべきだということを掲げている。両者

とも基準として用いている原理は、いわゆる人間の感じ方、すなわち、ヒューリスティックに基づいたものであり、それ自身に根拠を求めるることは非常に難しい。音楽的制約という観点から予測問題をコンピュータで扱おうとすると、音楽制約・知識の自己組織化の機構が十分にわかっていない現在、音楽要素の推移率等を人間が与え、それをもとに意外性とみるような方法しか取りえないであろう。ここでは、前者、すなわち、音楽的制約ができるだけ用いないという観点にたって、信号の規則性に基づいた構造予測法を考えてみることにする。



図4-9 譜例 久石譲：風の丘

図4-9に示す譜例で考えてみよう。ここでは、拍子、調などを知らされていないリスナーという条件で考えてみることにする。No.1段階では、予想は行なわれようがない。No.2の音に関しては、その時刻はわからないが、No.1と同じ音高、すなわち、B2が予測される。しかし、実際には、G3の音が、時刻2で響く。この段階で、No.1とNo.2の音の情報から、No.3の音が、時刻3で何らかの音が来るものだろうという予想が行なわれる。実際には、時刻5でB2が観測される。この段階で、No.3は、No.1と同じ高さの音であるから、No.4は、No.2と同じ高さの音が来るのではないかという予想がなされる。しかし、実際には、時刻6でF#3の音が響く。この段階でNo.1とNo.3の音が同じ音高であることに加え、同じ音価をもっているが確定するため、No.1とNo.2で一つのグループが構成され、No.3でその繰り返しが始まるという仮説が生成される。そうすると、No.4は本来、G3、音価が1拍のものではないかという予想がなされる。しかし、実際に音の高さに関してはF#3であるため、No.1とNo.2の構成要素のうち、2番目の音が1音下がるというモーションが検出される。次に、No.5の予想であるが、時刻9でB3の音が来るという予想がなされる。実際、その通りにNo.5の音が出現し、また、グループの2番目の音が1音下がるというモーションから、No.6の音に関しても予想通りであるため、この構成に関しての確信がもたれる。したがつ

て、7番目の音としては時刻13においてB3、8番目の音は、時刻14でD2が来るだろうという予想がなされる。実際には、時刻12においてD3の音が観測され、時刻13でC3の音が観測され、予想ができなくなってしまう。そして、時刻14の段階で、予想どおりのD2が来て初めて、最初のNo1, No2の音のグループに対する変形が現われたと理解される。音楽学的にも恐らくこれと同様の分析がなされるだろう。

このような分析を数理的な手法で扱うモデルとして、情報圧縮の原理に基づいた解析法を以下に述べる。具体的な原理としては、音楽は音のグループの引数付きの複製で構成されるとし、その操作において、パラメータが最小になるような音のグループを発見するという処理を行えばよい。この処理を最初から、音数を増やして行く過程で順次繰り返して行くことにより、構造を解析し、次の予測をするものとする。

例えば、図において、No.1からNo.4までの音に関しては、音高、音価、時刻の3つの情報があるので、計12個のパラメータが取られる。しかし、No.1とNo.2をまとめとし、それがどのように複製されるかという観点から整理すれば、No.1とNo.2に関する6個のパラメータ、および、その複製時刻とNo.2の音に対する音高の変化パターン変化の4つのパラメータの計10個で記述することができる。したがって、No.1とNo.2によるグループの核としての信頼度は高いものとして優先される。運動は持続するという考え方を使えば、この段階で、No.5の音の価はNo.1およびNo.3と同じであると予想される。

次に、No.1からNo.5までの段階では、No.1からNo.4に関してはNo.1とNo.2の複製が2個という記述になるが、No.5は個別の音となり、全部のパラメータの数は、 $6 + 2 + 2 + 3 = 13$ となる。あるいは、No.1が別で、No.2とNo.3が複製の元となる見方もあるが、パラメータ数は共に13で同じであるので、ここではどちらを支持してもよいとする。

その次に、No.1からNo.6までの音に関しては、No.1とNo.2に関する6個のパラメータ、および、その複製時刻とNo.2の音に対する音高の変化パターン変化の6つのパラメータの計12個で記述することができる。これは、No.1からNo.5の時よりも少ないパラメータ数になっているので、安定した（完結した）部分と読み取ることもできよう。しかも、複製時刻のパラメータ、音高の変化パターンに等差数列の傾向が発見されれば、初項、項差、項数という3つの記述することができる。例えば、上の例では、複製時刻に関しては、初項0, 4時刻毎に3回という2つのパラメータで記述ができる。このような方法を用いると、繰り返しが多ければ多いほど、情報の圧縮率が高まることがある。例えば、♪♪♪♪なる音の並びは♪♪♪♪という音の並びよりパラメータが多

く複雑とされるが、ある程度以上繰り返されるとパラメータ数は変わらず、逆に、
 ♪♪♪♪の方が等差数列の項数が倍のペースで大きくなる。項数の意味をどう見るかには問題もあるが、これは、聴覚上の特性ともよく一致したものとなっている。

上記の例では、No. 1 から No. 4 まで、No. 1 から No. 5 まで、No. 1 から No. 6 までというように徐々に大きくする様な形を例に示したが、人間の物理的な記憶容量の限界を考えると、解析の範囲をある程度限定するという方法も考慮していくことも今後の研究課題となろう。また、ここでは、音楽的制約というものは、できるだけ排除して考えたが、この解析法に加え、コード進行、あるいは、拍子などの音楽的制約条件によって、どのパラメータが必要なくなるかを探ることは非常に興味深い対象である。

4-5 検討

本章後半では、音楽聴取過程のうち感性状態の記述、および、抽出過程といった処理について述べてきた。伝播モデルとファクタモデルは、感性の自主性、記憶、発達、文化圏、状況による心情の変化などにおいて人間の感情や概念とは程遠いものであろうが、機械による感性情報のハンドリングの一つの可能性は示せたと考えている。外部から見かるかぎりこれらのモデルは先述のダイレクトモデルと大差はない。しかし、内部では、それぞれの音楽プリミティブに対する意味づけがなされており、例えば、感想文の出力においては自然な接続詞の挿入などが可能となっている。

伝播モデルとファクタモデルのうち実用という観点からはファクタモデルの方が適切である。ファクタモデルは、内部データが数少ない感性パラメータで記述される。インターフェースで扱うパラメータと機械内部で使用するパラメータを一一に対応させることができあり、作曲支援システム、解釈支援システムにおけるデータの検索、意志決定支援などの応用が期待される。

一方、予期モデルとしては、音楽的制約をできるだけ排除し、一般の信号における秩序と変化点に反応する手法を示している。音楽学や音楽認知の分野で示される方法の多くは心理過程や音楽的制約を必然として用いている。心理過程や音楽的性約に依存したモデルをコンピュータ上でインプリメントしようとすると、推移率やどちらがより結びつきやすいかなど、あまりにもヒューリスティックに与えざるを得ないデータや知識が多く、それに失敗するとモデル自身が意味をなさなくなってしまう。不十分ではあるが、ここで示した方法は、ヒューリスティックをほとんど要さないコンピュータ向きのモ

ルであると考えている。

4-6 まとめ

音楽の理解は、楽音の理解、楽音より高次の音楽理解に大別されるが、本章では、その後者を扱った。

1) 音楽を構成する要素（音楽プリミティブ）の認識解析手法

- ・拍子の抽出
- ・調性の抽出
- ・コード進行の抽出
- ・メロディの抽出

2) 感性情報の抽出・記述

- ・ダイレクトモデル
- ・伝播モデル
- ・ファクタモデル
- ・予期モデル

第5章 音楽解釈

この章では、音楽解釈過程、すなわち、演奏者が楽譜情報を解析し、その上で演奏表現を生成する過程のコンピュータ上での実現法について述べる。最初に、演奏の生成過程として、楽譜の認識、音楽構造の解析、表意記号の表現、音楽構造の表現について述べる。次に、実際の演奏からの演奏法の修得について述べる。楽譜情報を利用した効率的な演奏表現の抽出、表意記号、モチーフ、フレーズなどの表現ルールの発見、モチーフ抽出の戦略の学習などについて述べる。

5-1 自動演奏と音楽解釈

音楽に関する研究領域でA I及び認知科学に関連するものには、自動作曲・作曲支援、音楽解釈のモデリング、音楽聴取過程のモデリングなどがある。

このうち、2番目の演奏者の音楽解釈のモデリングは、音楽情報処理のなかでも主要なテーマとなっている自動演奏、自動伴奏の今後の研究の根幹となるものである。

自動演奏は17世紀初頭のオートマタと呼ばれる発音機構のある人形を初めとし、オルゴール、ロールピアノなどの演奏装置にまで遡ることができる。計算機が演奏自体に応用されるようになったのは、ミニ・コンピュータが普及し始めたころである。そして、現在に至っては、マイクロコンピュータの普及とMIDI規格により、アマチュアレベルからプロフェショナルまで幅広く計算機を利用した自動演奏が行なわれている。機械によるシーケンサは、正確無比な演奏、人間で到底できないような速いパッセージの表現を可能としたため、コンピュータミュージックと言えば、まさに機械的な音楽であるというイメージと結び付けられることもあった。最近では、これとは逆に「人間的な演奏はどのようにしたらできるのか」ということが話題にあがっており、情緒溢れる演奏データをインプットする作業、いわゆる「打ち込み」という、今までの音楽制作形態ではなかったプロフェッショナルも登場している。このような音楽における人間性への回帰とA I的な興味が本章で扱う音楽解釈システムの背景となっている。

一方、エレクトロニクス技術の発達は、従来夢の世界にあった自動伴奏という具体的なアプリケーションを可能にしつつある。現在、かなりのレベルで演奏者のテンポに追従するシステムも紹介されているが、処理的には単純な信号処理が実現されているに過ぎない。

ず、欲を言うと、人間の伴奏者のように「気を利かす」ことは出来ない。より、発展した形としては、やはりシステムがある種の解釈機能を持っている必要があるだろう。

解釈的処理を実現した研究は2、3見られるが、最も処理過程に留意した研究としては階層的な表現が楽曲の演奏を決定するという考え方の元にシステムを作ったClynesの研究ができることができる[Clynes 1984]。例えば、 $3/4$ 拍子の楽曲のある4小節が48個の16分音符から成り立っている場合、それぞれ4つの小節に対する音量・支配時間の割当、1小節内での各1拍目、2拍目、3拍目の大きさ・支配時間の割当、1拍内のそれぞれの4つの16分音符の音量・支配時間を比で表現し、その積算をもって演奏データを作成している。その他に、メロディに対する表現、ビブラートのかけ方に対する演奏ルールが議論されており、考え方としては本章で述べる解釈システムMIS(Music Interpretation System)[Katayose 1990 b]と一致するものである。

国内の研究としては、PSYCHE[Tomizuka 1988]、MUSE[Taguti 1989]、そして本章の主題であるMISなどがある。PSYCHEは、予め、さまざまなフレーズの演奏法をデータベースとして登録しておき、自動演奏を行おうとする楽曲の同等あるいは類似のフレーズに対応する演奏法をデータベースから検索し、つなぎ合わせることにより演奏データを作成している。一方、MUSEは演奏記述言語であり、それ自体は解釈機構を目指したものではないが、階層的な演奏パラメータの記述が可能となっており、ワルツの伴奏表現の数理的な解析などに有効に使われている。

5-2 演奏解釈システムMIS

音楽分野において解釈とは、第一義的には、演奏家がある作品に自分の音楽的見解を確立し、作品を主体的に把握し、演奏に反映することであると定義されている[音楽の友社 1975]。解釈に考慮されるべき要因としては、その歴史的意味、曲風、形式構造を始め、速度強弱、各種の表現法などあらゆる作曲的および演奏的問題がある。音楽美学においても、演奏解釈は主観派、客観派をはじめとして様々な議論がかわされてきたが、心的要素の位置づけなど哲学に包含されるものが多く、その具体的なモデルとしては現在進行中の領域である。しかし、解釈過程においては、機能的には、次に示す2つの処理が、実現されるものでなければならないだろう。

1. 楽譜を入力に対して、情緒溢れる演奏を行う
2. 演奏法を自動的に学習する。

音楽解釈システム "MIS" はこのような機能の実現を目指して構築されたものである。 MIS の概要を図 5-1 に示す。

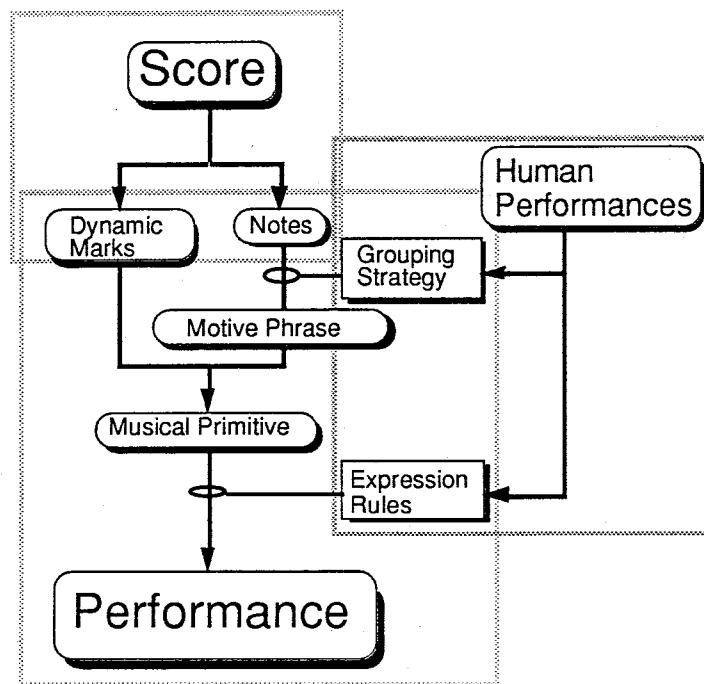


図 5-1 音楽解釈システム MIS

図 5-1 に示すように、 MIS は楽譜認識過程、演奏生成過程、演奏ルール学習過程の 3 つのステージから成り立っている。このシステムの実現においては、解釈に使われる知識はルールで表現され、全演奏データは各ルールによって生成されたそれぞれの表現要素の和によって決定されるという前提を用いている。本来解釈の対象とすべき音楽はさまざまであるが、ここではピアノを対象に実験を行なっている。解釈の対象となる要素はモチーフ、フレーズの表現、表意記号の表現、楽曲の種類（背景）、そして、感性情報の表現である。このうち、表意記号、楽曲の種類（背景）は楽譜から直接的に与えられる情報であり、モチーフ、フレーズは楽譜を解析してえられる情報である。感性情報は、例えば、ピアノの先生が生徒に演奏を教える時に「この部分は重たく」、「この部分は弾んで」というように、人間がシステムに対し直接的に示唆する情報である。

MIS によって生成される演奏のグラフ表現を図 5-2 に示す。

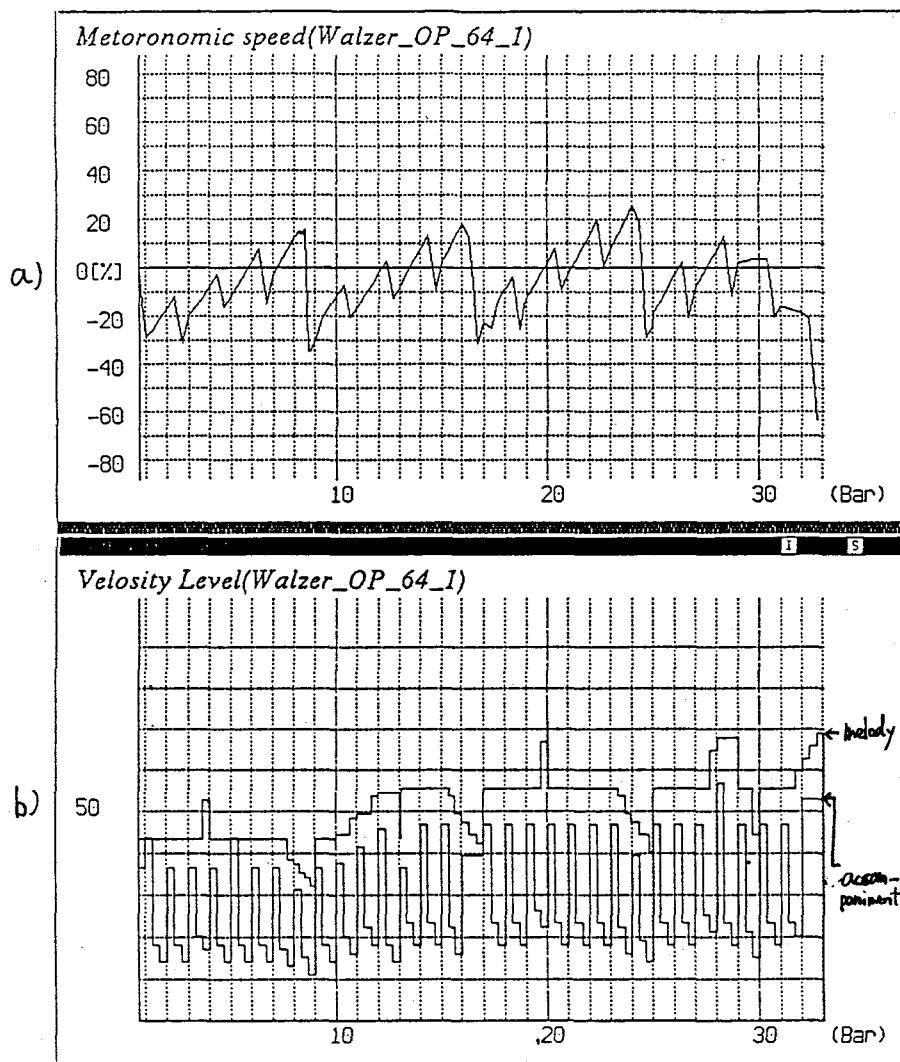


図5-2 MISにより生成される演奏例（ショパン作曲：小犬のワルツ）

a) メトロノーム速度分布 b) ベロシティ（楽譜は付録5参照）

図5-2 a)において、メトロノーム速度とは、曲の早さを示す指標である。詳しくは付録6で述べる。

5-2-1 楽譜認識過程

楽譜認識過程では、画像データ（楽譜）が読み込まれ、解釈過程の入力データとなる音符データ、表意記号のシンボル情報に変換される[Kato 1990]。楽譜はイメージスキャナを通して取り込まれ、前処理として、5線の太さ、間隔などを検出したうえで、小節単位に分割される。小節単位に分割した画像に対して一種の“黒板モデル”的な手法で下位レベル（画像レベル）から上位レベル（意味レベル）へ仮説の生成・検定を繰り返しながら1小節の長さは一定であるというゴールに向かって認識が行われてゆく。また、画像中の各図形は各シンボルの認識が進むにしたがって消去されてゆく。最終的には認識対象外の図形のみが残されるため、認識の対象シンボルを新たに付加していくことは比較的容易である（図5-3）。このシステムによる認識結果を表5-1に示す。

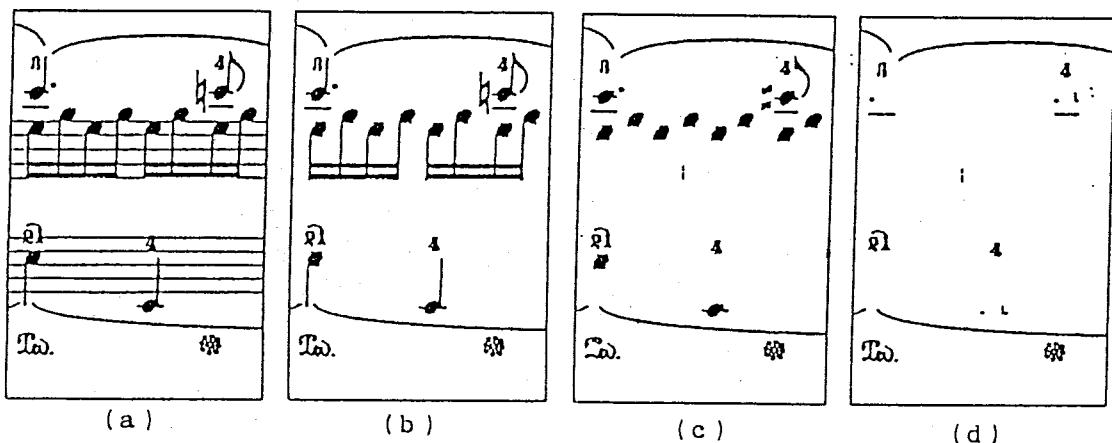


図5-3 認識過程

表5-1 認識結果

Name of piece	Recognition rate
For Elise (L.V.Beaethoven Op.173)	95.6%
Turkish March (W.A.Mozart)	91.5%
Etude (F.Chopin Op.10 No.3)	87.1%
Sonata Pathetic 2nd movement (L.V.Beaethoven Op.13)	83.3%

5-2-2 演奏生成過程[Fukuoka 1990]

演奏生成過程は音楽解釈の中核となる過程である。表現の対象となるのはクレシェンド、スタカートなど楽譜に明示的に記述された表意記号、および音列から解析してえられるフレーズ、モチーフなどである。ここでは表意記号およびモチーフ、フレーズなど演奏表現のもととなる対象を音楽プリミティブと呼んでいる。音楽プリミティブに対し表現ルールを適用することによって実際の演奏情報が生成される。全体的な表現は各表現の和として生成される。グルーピング戦略はモチーフを抽出するための戦略である。特定の音列が音楽中で現われたら、それを一つのグループであると優先的にとらえることにより、それをまとまり感がつくように表現を行なう。対象となる音楽に適合する音列がないときには、拍子を満足するグループが選択される。ここではグルーピング戦略と表現ルールをまとめて演奏ルールとしている。

5-2-3 演奏ルール学習過程

楽譜認識過程、演奏情報生成過程の二つの流れは生成型のエキスパートシステムとして位置づけることが出来る。MISに登録したルールにより生成したデータに対し、聴取実験を行なうことにより、人間が演奏ルールを精製してゆくという方法も取りうる。しかし、MISの他の解釈システムともっとも異なる点として、演奏例を与えることによって自動的に演奏ルールを自立的に学習する機能を有している。この部分はMISの中核でもあるため、5-6節から5-8節までにわたって詳しく説明する。

5-3 データ表現とルール

5-3-1 データ表現

読み込まれた楽譜データはscr形式で記述されている。演奏情報としてはnote形式を用いている。解釈機構における処理はscr形式からnote形式へのデータ変換過程を考えることができる。データ例を図5-4に示す。

```

composer Chopin           D4 48 12 (right)
songname Valse_op.69      D3 48 24 (left)
key G                      B2 48 24 (left)
tempo quater 152          F#2 48 24 (left,pedal_off)
==<1>==                  B3 60 12 (right)
F#4 0 24 (right,acc, slur,slur) ==<3>==
R 0 24 (left)              A#3 0 48 (right,slur_end)
==<2>==                  C#2 0 24 (left, pedal)
F#4 0 12 (right,slur_end, p) E3 24 24 (left)
B1 0 24 (left, pedal)     C#3 24 24 (left)
G4 12 12 (right)          F#2 24 24 (left)
F#4 24 12 (right)         F#4 48 24 (right, acc, slur, slur)
D3 24 24 (left)           E3 48 24 (left)
B2 24 24 (left)           C#3 48 24 (left)
F#2 24 24 (left)          F#2 48 24 (left,pedal_off)
C#4 36 12 (right)         ==<0>==


```

scr DATA

```

Header{
Beat 4/4
EXP ON
Tempo 0.5<----- 4分音符のテンポ
}

----- 0 /* 小節線の位置 書かなくてもよい */
0 ( G3 +0.0 1.0 -0.1 60 ) (+0.1 A2 1.0 -0.1 46 ) (+0 D2 0.5 -0.05 52 )
Tempo 0.8 <---- テンポの変更
0.5 (-0.1 F3 0.5 +0 24)
Tempo 0.7 <---- テンポの変更
2 (+0.2 B3 1 +0 23)
Tempo 0.5 <---- テンポの変更
3 (+0.05 F4 2 +0 78 ) (D3 +0.1 2.0 -0.1 45 )
Rest 0.2 <---- 休止区間の挿入
----- 1 /* 小節線の位置 */

.
.
.

END

```

Note DATA

図5-4 データ形式

note形式において、時間の制御に関しては、任意時間におけるテンポのコントロール、音の開始時間、持続時間の変化、および“間”的挿入が可能な形式となっている。実際には絶体的な音の立ち上がり時刻と持続時間を記述するだけで十分であるが、人間を主体に考えると

1. 拍に基づいてテンポをとっているときの表現（フレーズなど）
2. 1拍以下の長さの微妙な表現（乗り、モチーフなど）
3. 拍では換算できない休止区間

に対する処理は、手拍子のオーダーで制御されるものかどうかという点で分離されるものである。ここでは、演奏ルールが人間にわかりやすいようにするために、フレーズなど全体的に処理されるものに関してはテンポの変化と休止区間、モチーフの表現に関しては相対的な時間比をルールで記述するようにしている。note形式における音の大きさはMIDIでいうベロシティ値を用いている。また、今回は対象をピアノに限定したため、ペダルのみをコントロール信号として送信している。

5-3-2 データの内部表現

ルールの条件部に使われるデータの内部表現は図5-5に示すようにノート情報および属性情報をノードとするリスト表現で記述されている。

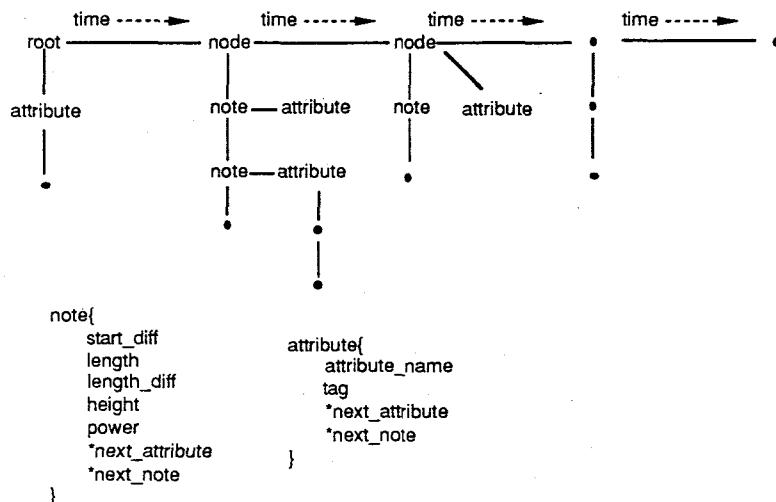


図 5-5 楽譜データの内部表現

ここで属性情報とは、名前と識別番号のセットから成り立ち、動的に操作できる環境が作成されている。属性情報には楽譜から得られるシンボルおよびモチーフ、フレーズなどの構造解析処理から得られた結果が記述される。ここで楽譜から得られるシンボルとは表意記号、連符、ピアノであれば音部などの楽譜情報から得られる構造上必要と思われるすべての情報である。

5-3-3 ルールの働き方

表現ルールは属性記号に対するプロダクションルールとして表現される。フレーズに全体にかかるテンポのコントロールとアッヂェレランド、リタルランドなどの標語に関しては予め定義した関数（後述）にパラメータを与えるという形で表現され、ノートデータの予約語Tempoを使って演奏データに変換される。その他の表現に関しては直接的に各音各音のスタート時間、持続時間、ベロシティ値を変化させることにより表現する。

モチーフの表現、スタカートなど予約語Tempoを用いない表現ルールは“比”で表現される。ある音列に対し、複数の表現ルールがマッチするときには比の積によって演奏データが生成される。この様子を図5-6に示す。なお、属性情報に関するネーミングおよび表現ルールについては隨時追加できるようになっており、聴取実験の過程を通じた演奏ルールの獲得が可能となっている。

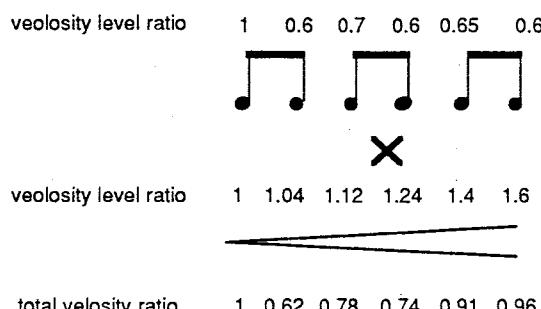


図5-6 演奏データの生成

5-4 表現ルール

演奏に表情を付けるためには、まず、楽譜上に表記されている表意記号の表現を上手に行なってやるというのが第一歩であろう。そして、それ以上に大事なことは、対象曲の構造を理解して、明示的には書かれていない音楽の表現を付加することである。ここでは、表意記号、そして、構造が理解された段階で実際に演奏を生成するための知識を表現ルールとしている。表現ルールの抽出に相当する研究は音楽学あるいは心理学分野において早くから研究が行なわれている。この分野における当初の興味は楽譜と人間の（名）演奏の違いを計測するという点にあり、最初の研究は1920年代に始められている。Seashoreらは声楽、バイオリンのピッチ変動やビブラートの深さ、さらにはピアノを含

めて演奏速度のテンポのずれの実測を行なった。そのうえで、楽曲の繰り返し時に見られる表現の類似性などを根拠に解釈の一意性について論じた[Seashore 1938]。それ以来、いくつかの研究結果が報告されている。Gabrielssonらは単旋律を対象に連続する音価の表現を中心として演奏ルールを実験的に見いだそうとした。一つの例をあげると2分音符と4分音符一つの組み合わせの実音価の比は2:1ではなく1.60~1.85:1であるというような実験結果が述べられている[Gabrielsson 1985]。さらに明確にルールの抽出という点に注目をおいた研究としてはSundbergらの研究がある[Frydon 1984]。彼らはルールベースで演奏を制御するシステムを構築し、アーリシス・バイ・シンセシスの手法でコードの変化(5度圏のトニック距離)、フレーズ・サブフレーズの境界、コードから見たメロディの個々の音の関係に対する表現ルールの抽出を行っている。

MISにおいては、このようにルール化された表現法は簡単にインプリメントできる。また、その演奏を通じてそのルールの評価が可能である。次に、MISにおける特徴的な表現ルールとその展開法を示す。

5-4-1 フレージング

フレージングの最も基本的な概念は、音楽的まとまりを表現することである。この「音楽的まとまり」は、同じ、あるいは、類似した構造に対し同様な表現を施すことによって秩序付けられると考えることができる。音楽的まとまりを構成する要素は本来様々な要素があるが、ここでは最も中心的と考えられるメロディを取り上げる。また、「音楽的まとまり」表現がサブフレーズ、フレーズというように階層的に重なり合わあってゆくことにより対局的な演奏が生成されると考える。ワルツを対象にした一般的な表現の例を図5-7に示す。

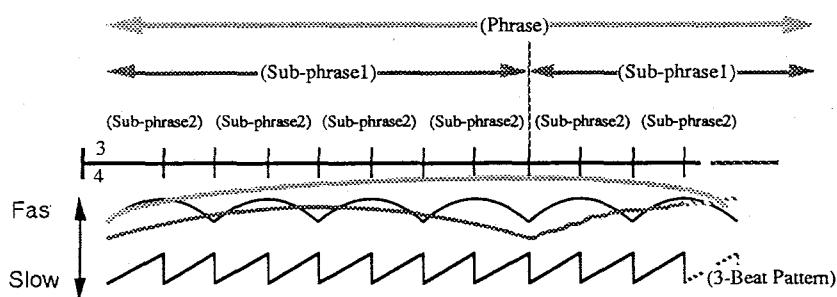


図5-7 フレーズの階層的な表現

この例においてはワルツ特有の表現法が左手のパートに表されている。右手の表現は各階層における表現の和で与えられる。実際の演奏における演奏（ショパン作曲「ワルツ第10番」）からテンポを解析した結果、および抽出したパラメータから再合成したテンポデータを図5-8に示す。

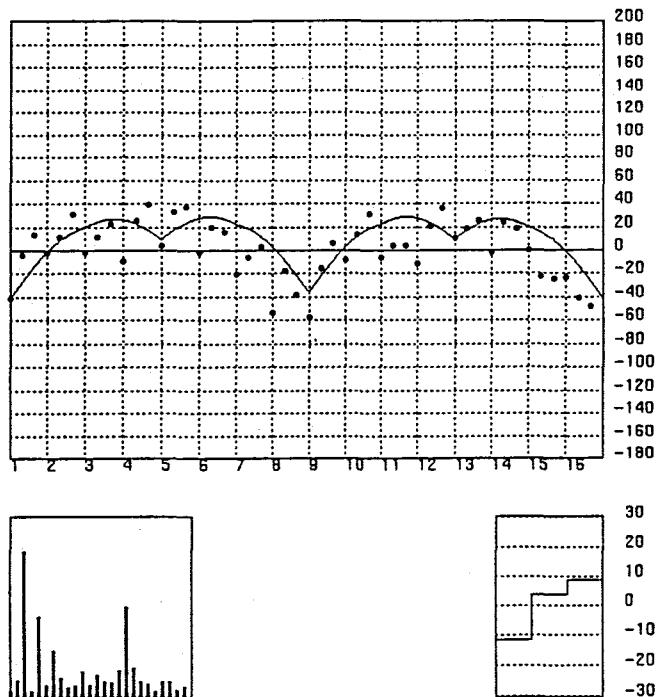


図5-8 パラメータの解析例 ワルツ第10番（中間部第1小節から第16小節まで）
上：演奏速度分布（ドット：実データ、曲線：抽出したモデルのフィッティング）
下左：スペクトル分布、横軸は周波数 下右：3拍子パターン分布

この図におけるテンポには局所メトロノーム速度と標準速度からの偏差分布を用いている。グラフ中の点列を見てもわかるように実際の演奏分布には一見揺らぎともとれる拍数以下のモチーフの表現がある。下左図はメトロノーム速度分布を一次元の波形と考え、スペクトル分布を求めた結果である。この結果は周波数が2、すなわち16小節／2 (=8小節)に対するまとまり付けが最も強く、続いて、16小節／4 (=4小節)と16小節／16 (=1小節)にまとまり付けが付けられていることを示している。上図の実線は4小節、8小節の周期を持つ成分を逆変換したものであり、このデータは階

層的なフレーズの表現の重ね合わせで全体的なテンポの変化がほぼ正しく付けられることを示すものとなっている。また、このような組み合わせが、演奏における起承転結を構成すると考えることができる。

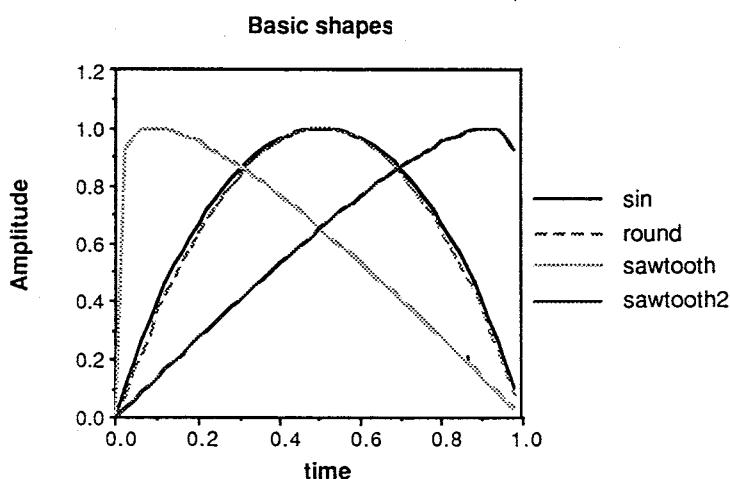
フレーズ表現に適用されうるテンポ表現は様々なものが考えられるが、基本的には、「まとまり（フレーズ）の開始時終了時は、ゆっくりと演奏する」、「ゆっくりとスタートして段々早くなって行く」、「最初は早く段々遅くなって行く」の3種類に分類される。ここでは、この3パターンを表現する関数として統計学上のベータ関数の積分の独立変数をとったベータ関数A(t)を利用する。

$$A(t) = \frac{G}{N} \left(\frac{t}{T} \right)^{p_1} \left[1 - \frac{t}{T} \right]^{p_2} \quad 0 \leq t \leq T$$

ここで、Gは振幅、Tは変化を付ける区間の長さ（時間）である。Nは関数を正規化するために使われ、次式で与えられる。

$$N = \frac{p_1^{p_1} p_2^{p_2}}{(p_1 + p_2)^{p_1 + p_2}} \quad p_1, p_2 > 0$$

実際にp1とp2を変数、G = 1, T = 1とした時の関数を図5-9に示す。



名前	p1	p2
sin	0.89	0.89
round	1.0	1.0
sawtooth1	0.001	1.0
sawtooth2	1.0	0.001

図5-9 ベータ関数

どのフレーズに対してどのようなパラメータを取るかという問題については、デフォルトルールを用意しておき、特別な表現や演奏を規定する種類が規定されているときにそのルールを適用するという形式を取るものとする。デフォルトルールは2小節フレーズに対しては、「まとまり（フレーズ）の中央部でもっとも速くなる形」というパラメータを用いる。ベータ関数の例では、 $p1 = p2 = 0.89$ 、Gについては、それぞれ16小節、8小節、2小節にあてはまるパラメータとして $G_{16}=5.2 \sim 14.4$, $G_8=21.6 \sim 47.2$, $G_2=15.0 \sim 25.6$ となることが多くの演奏例より確かめられている[Takami 1989a]。

5-4-2 モチーフの表現

フレーズの表現では、テンポというパラメータをコントロールすることにより、緩急法を表現したが、モチーフに対しては、その部分の音のスタート時間、長さ、ペロシティを変化させることによって表現を行なう。したがって、ルールベースとしては、モチーフ群とそれに対応する個々の音の演奏パラメータが与えられる。例えば、を強調するという表現ルールは、二つの音の長さの比率を4:1に、ペロシティ比1.5:1にするという形で与えられる。

図5-10にワルツに特有のモチーフおよび演奏パターンの強弱モデルを示す。この左において左のモチーフは、1小節における連続する8部音符の演奏法で、強拍、弱拍をモデル化したものとなっている。なお、このモデルは、拍子に対して矛盾なくグレーピングされたモチーフに対して適用されるものである。モチーフ、フレーズの認識については、5-5節で述べる。

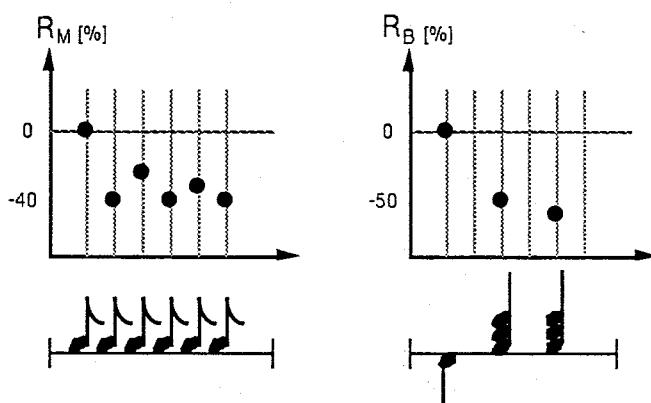


図5-10 モチーフ、伴奏パターンの強弱モデル

5-4-3 モチーフ構造に働くルール

楽曲においては、しばしば（フレーズ以下の）短い音列やモチーフが連續した繰り返しとして扱われることがあり、それが演奏に影響を与えていていることが多い。このような構造的表現を実現するために、音列にグループとしてのタグを持たせ、そのタグの並び具合自身に演奏ルールを与えることを許している。例えば、図5-11の例で音列グループXが4度繰り返せば、X1, X2, X3, X4の長さを指数的に短く（テンポをあげて）、ペロシティを上げるというようなルールを記述することができる。

構造に働くルールとしてもう一つ重要なものに、主旋律と伴奏等の音量のバランスがある。scrファイルに旋律を示す属性を付加することにより、演奏データの生成という問題は解決されるが、条件節に対しては難しい問題が残る。ピアノ楽譜に限定した場合は、音部により旋律と伴奏部とがかなり明確に分離が行なわれているため、ト音部を旋律とすることにより、かなりの部分は解決される。しかし、根本的な解決を見るためにはポリフォニー中からメロディパート、サブメロディパートを抽出する機構が必要である。

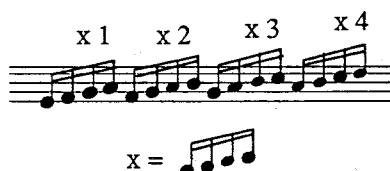


図5-11 音列（モチーフ）の繰り返しに対するルール

5-4-4 曲の種類が規定するモチーフ表現

曲の構造が同じでもその種類が異なれば演奏法を変える必要がある。例えば、同じ3拍子系の曲でも、ワルツとマズルカでは本来異なったリズム表現を持つものとされる。ある楽曲がワルツであるかマズルカであるかを音楽構造だけから判断することはほとんど不可能である。MISでは楽曲の種類は、作曲者がその曲の弾き方を強く規定するため定めたものとして、楽曲の種類に応じて表現ルールを選択するものとしている。

図5-12に示す例はジャズのスタンダードの「Take the A Train」である。図中で、旋律の6小節目のメロディー列（8分音符）は決して均等の長さで演奏されるものではない。

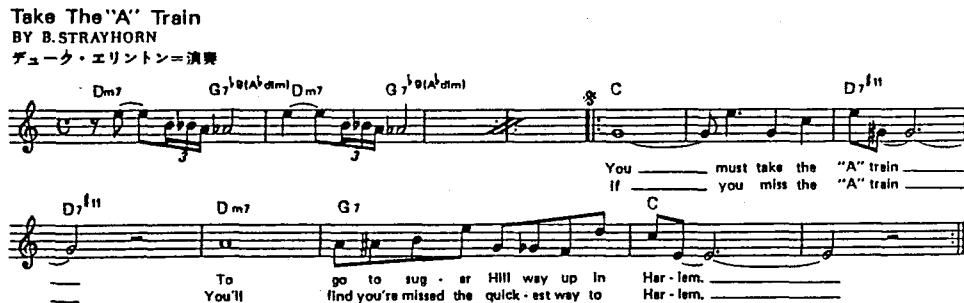


図5-12 Take the A Train

通常、ジャズにはスイング感がつきもので、ウラ拍（2, 4, 6, 8拍目）はそれぞれ、次のアタマ拍（3, 5, 7, 次小節の1拍目）に引っ掛けるように演奏されるべきである。これは対象がジャズであるという種類が明示的に与えられたときに使われるルールである。

5-4-5 表意記号の表現

上記のフレーズ、モチーフの表現に比べて、表意記号はどのような表現を行なえばよいかが明示的に与えられている。そのため、予め、表現の関数を設定しておき、パラメータで表現を変えることができる。現在、ペダル、スタカート、テヌート、リタルランド、アッチャレランド、リタルランド, *f ff mf p pp mp*, クレシェンド、デクレシェンドなどの表現法が登録されている。強弱表現の例を図5-13に示す。

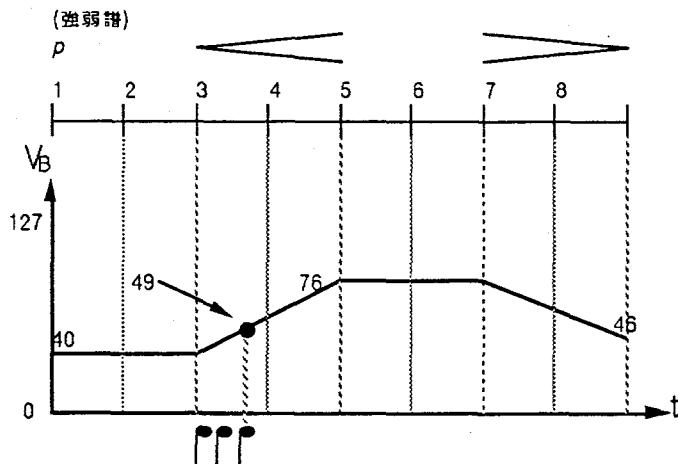


図5-13 強弱表現の例、縦軸：ベロシティ、横軸：時間

5-4-6 感性情報の表現

感性情報の表現には演奏者の解釈要素が強く出される。この要素の完全な形は、本来言葉で表せるものではなかろうが、優しくとか、重たくというような言葉がどのような表現に結びついているかを探ることも重要である。現在、「弾んで」と「重たく」、「おだやかに」と「激しく」が感性情報にむすびついた表現ルールとして用いることができる。「弾んで」と「重たく」については、音のオンタイムという物理的な要素に展開できるものである。「おだやかに」と「激しく」については、今まで述べてきたようなパラメータを強調するという形で実現される。

5-5 楽曲の構造解析とグルーピング・ストラテジー

演奏情報を生成するための大きな要因としてモチーフ、フレーズなどの表現がある。本解釈システムの大きな特徴の一つは自立的に構造解析を行ったうえで、モチーフ・フレーズなどの演奏法を抽出することである。

ほとんどの楽曲においては、2小節、4小節など2の階乗小節でフレーズが構成されるが、音楽の最小構成単位であるモチーフは、その長さ、位置はトップダウン的に与えられるものではない。また、フレーズも弱起などの楽曲においては小節線がフレーズの単位となることはないため、ボトムアップ的な構造解析が必要となる。図5-14に弱起の曲の構造を示す。

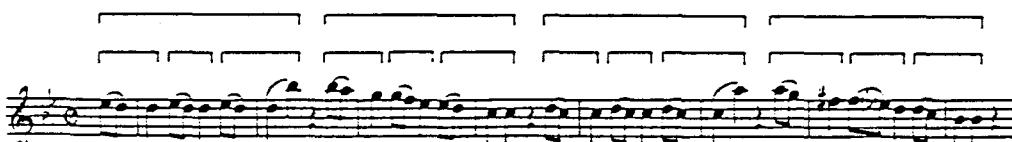


図5-14 モーツアルト：交響曲 No.40 K.550 第1楽章

音楽構造を決める要素としては、リズム、メロディ、ハーモニー、サウンドなどの様々な要素がある。また、メロディ中のモチーフ、フレーズを解析するだけでも、音の高

さ情報と長さ情報を相互的に考慮していく必要があるが、ここでは問題を簡単化するためにメロディに注目する。上記の要素のうち、リズムは、特に音楽的構造と密接な関係があり、構造の把握とリズムの認識はそれぞれの相互作用によってえられるといつても過言ではない。リズムに関しては音楽分野でさまざまな理論や考察がなされているが、その最も基本となるのは繰り返しパターンの知覚であると考えられる。ここでは、メロディのメロディのアップ・ダウン（メロディ構成音の差分）、サーフェイス・リズム（メロディのビート）の繰り返しパターンに基づいたモチーフ、フレーズの解析法を示す。概略は高さ情報と長さ情報から繰り返しパターンを求め、その繰り返し周期分を置いた同等、あるいは、類似の音列を領域成長の手法で発見するというものである。ここで、モチーフは対象となる曲において少なくとも2回は繰り返されるという前提を用いている。ヒューリステックスとしてマイヤー的基準が利用される。これは、同じモチーフが連続して何回も繰り返して出現するような楽曲においてモチーフ、フレーズのスタート位置を発見するためと、探索の効率化のために使われるものである。グルーピングのアルゴリズムを以下に示す。

Step 1 メロディ中の最も短い音の長さを1単位として、図5-15に示すような音のアップダウン、サーフィスリズムの配列を作る。音のアップダウンに関しては、前後の音高の差分によって与える。サーフェスリズムに関しては、音がオンになる部分を1、それ以外を0としている。ここでは、メロディを対象としているために、配列の要素は1か0になっているが、ポリフォニー（伴奏つき）の場合は、そのタイミングでの音数を要素とすればよい。

Step 2 Step 1で求めた配列 $x(t)$ に対し、次に示す式で $y(\tau)$ を求める。

$$y(\tau) = \sum_{t=0}^T \text{abs}(x(t) - x(t+\tau))$$

$y(\tau)$ が小さければ小さいほど時間 τ における繰り返しが強いと考えることができる。 $y(\tau)$ が小さいものから順にStep 3を行なってゆく。

Step 3 Step 2で求めた繰り返し周期 τ において同じ形状あるいは類似した形状をもつ音列を抽出する。具体的には、時間 τ または τ の整数倍において連続する2音が同じ形状でかつ複数個ある時、多数のループを形成するものから、その音列の形状を $\text{group}_k(t, 2)$ （種）を作成してゆく。この時に、同じモチーフが連続して何回も繰り返して出現する場合、すなわち、繰り返し周期 τ より小さい区間のずれでモチーフ候補群が複数個あるときには、「最初に来た音を優先する」「長い音

がフレーズのおわりの音である。」などのヒューリスティックルールが使われる。
 $\text{group}_k(t, p)$ において k はシードのタグ, t は始点 p は連続する音の数とする。 $\text{group}_k(t, p) = \text{group}_k(t+n\tau, p) \text{ and } \text{note}(t+p+1) \doteq \text{note}(t+n\tau+p+1)$, であれば p に $p+1$ を代入し, group_k を成長させる。この作業を順次繰り返して行く過程で group_k の数が変われば、すなわち、モチーフの変形が発見されればその時点で派生グループ group_k を作成してゆく。 group の成長は同じ形の音列がなくなるまで続けられる。

Step 4 τ を変えて Step 3 を繰り返す。

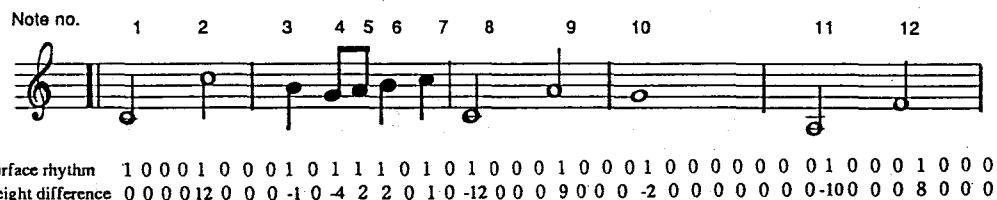


図5-15 音の配列

Step 3 で用いるヒューリスティック基準はすべての対象曲に対して当てはまるものではない。例えば、ボサノバなどにおいては、モチーフの冒頭に強いシンコペーションがあるため、フレーズの最後に長い音符が使われるというヒューリスティックを使うと失敗を起こしてしまう。しかし、人間にとっても楽譜あるいは伴奏付きのデータではなく単なる音の並びでシンコペーションの強い楽曲を与えたときには、恐らく構造を理解することはできないだろう。

伴奏自身がリズム上の制約を作り、それがメロディの構造を作っているということは、メロディと伴奏付きのメロディではどちらが記憶しやすいかという実験で簡単に確かめることができる。また、Step 3において \doteq が使われているのが類似性の処理に相当するものである。類似性の尺度を決定することは非常に難しいが、ここでは、成長したグループの音数に応じて予想される音との違いを相殺するという処理を実現している。上記の方法による分析過程を図5-16に示す。このデータは、拍子に非常に密接なグループングが行なわれている。

一方、図5-17に示す譜例に対しては、繰り返しパターンが1, 3, 2, 4(拍)の順で抽出される。このメロディは3拍子系であることを考慮すると3(拍)が上位で検出されることは妥当であるが、2(拍)の倍数で繰り返しパターンが検出されていることが興味深い。

Andante très doux
note No. ⇒ 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

Music by KEN Muramatsu
C 1994 by CBS/Sony Songs
2a 2b 2c 2d 2e

(a)

id = 1
 $(0, 3) (6, 9) (15, 18) (20, 23) (26, 29) (46, 49)$
 $(55, 58) (60, 63) (66, 69) (77, 80)$

id = 6
 $(2, 9) (11, 18) (22, 29) (42, 49)$
 $(51, 58) (62, 69) (73, 80)$

id = 2 extension of id = 1
 $(0, 5) (20, 25) (60, 65) (66, 71)$

id = 7 extension of id = 6
 $(2, 10) (22, 30) (42, 50)$

id = 3 extension of id = 2
 $(0, 9) (20, 29) (60, 69)$

id = 8 extension of id = 7
 $(2, 29) (42, 69)$

id = 4 extension of id = 3
 $(0, 12) (20, 32)$

id = 9 extension of id = 6
 $(11, 19) (51, 59) (73, 81)$

id = 5 extension of id = 1
 $(15, 19) (55, 59) (77, 81)$

id = 10 extension of id = 9
 $(11, 21) (51, 61)$

(b)

図5-16 グルーピングの過程

括弧は（グループの始まりのノート番号、グループの終わりのノート番号）を示す。idはグループの番号を示している。

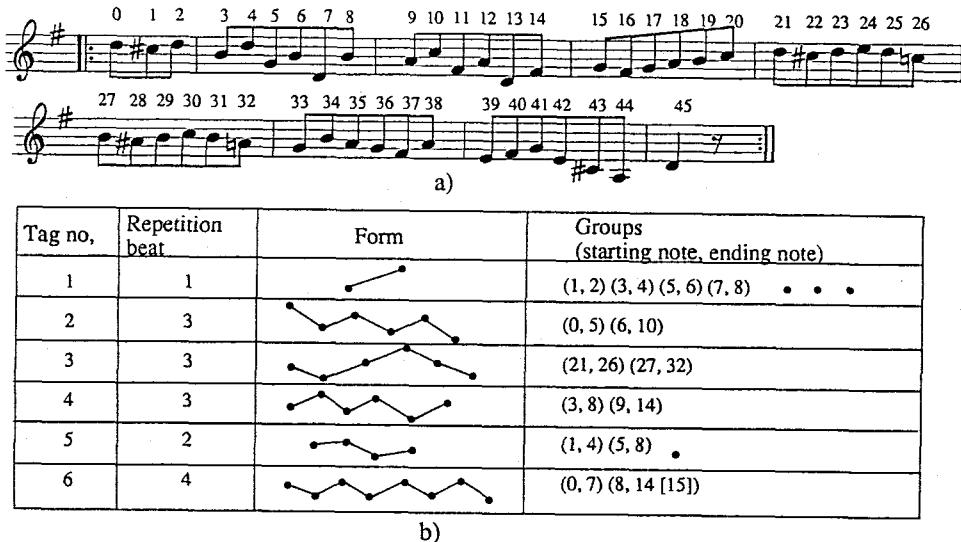


図5-17 拍子と矛盾したグルーピング

さらに、グループとしては、(0, 7) が抽出されており、これは(0, 5)と(6, 10)というグルーピングと矛盾するものとなっている。これに対し、作曲者の意図する音楽的構造は一意であるということから、トップダウン的に構造を決定するべきだという考え方を取る人もいる。確かに、作曲者が一つの作品を作るときにはその作品に対する構成上の意図があるのはわかるが、それでは最初にどのような構成を設定したのかという問題になると、実際問題作曲者のみぞ知るということになろう。仮に、人間が楽譜を見て“直感的に”理解できたとしてもその“直感的に”という作業の間にいわゆるボトムアップ的な見方で作曲者の構成に対する仮説を検定し、その仮説を説明する段階においてトップダウン的な処理が取られると考えたほうがよい。したがって、どのように使うかは別としても、構造上に多重の構造的特徴があれば認識しておく必要はある。実際、拍子に対しては逸脱しているにもかかわらず、音列構造上のグルーピングが表現が非常に心地よい演奏を産み出しているという例も紹介されている[Murao 1989]。さて、その曲が持つ多重のグルーピングが解析されることは示したが、実際の演奏を行なうときにはどのグルーピングを取るかを決定して必要がある。ここでは、どのモチーフグループを優先するかを決定する基準をグルーピングストラテジーと呼んでいる。この基準は作曲者というより、むしろ、演奏者に固有のものである。グルーピングストラテジーが規定されていない段階では、拍子に適合するグループが演奏を決定するものとして用いられる。

5-6 演奏分析[Katayose 1990b]

本章冒頭でも述べたように情緒ある演奏を生成するシステムの研究はほかにも2, 3見られるが、そこで使われる演奏ルールは人間が与えるというものになっている。Sundbergらのアプローチにしても人間がルールの仮説を立て聴取実験を行うというものであり、システムが自立的にルールの抽出を行なうというものではない。本節以降ではコンピュータが情緒溢れる演奏を聞いて、それから自立的に演奏ルールを抽出するという機構について述べる。抽出・学習される対象は表現ルールとグルーピングストラテジーである。グルーピングストラテジーはどのような音列がモチーフ、フレーズであるかを決定するために使われる戦略である。表現ルールは表意記号、音楽の構造を具体的な表現に展開するためのルールである。

MISにおける演奏ルール抽出の大きな特徴はモチーフの表現の学習に見られる。モチーフの表現の学習においては何がモチーフであるかをシステムが理解している必要がある。簡単には、具体的に対象となるモチーフを登録するという方法も考えられるが、これでは真の意味での学習とはいえない。MISにおいては、構造解析の手法をシステムに与え、考えられるモチーフ候補のうち、特徴的な表現のあるものについて表現ルールを抽出するという方法がとられる。この過程は、まず顕著な表現を発見してから、その理由を楽曲の構造に見いだすという音楽学者の研究スタイルにも相当するものである。

MISは実際の演奏から演奏ルールを抽出するという立場を取っており、自動的に音の立ち上がり時間、音の強さなどの情報を得る必要がある。これらの演奏情報は楽譜情報を手掛かりに楽音の時間周波数マップから抽出される。本節では、先ず、音響信号から、微妙な表現を抽出する機構について述べる。

5-6-1 音楽表現の抽出

この処理の目的は実際の演奏から各音の長さ、大きさを抽出することである。これは採譜処理の対象としてとらえることもできるが、テンポの速い楽曲やダイナミックレンジの大きい演奏に対しては採譜処理では対処できないことがある。ここでは楽譜をモデルとして導入することにより、自動採譜が扱えないレベルでの音の扱われ方の抽出を行う処理について述べる。これはちょうど楽譜を頭に描きながら注意深く演奏を聞く過程に相当するものである。処理の概念図を図5-18に示す。中心的な作業は楽譜情報の入力、音響信号の周波数分析、それを画像化した時間周波数マップと楽譜との対応探索

である。ここではまず、音の立ち上がり時刻の対応を探索し、その後で、他の要素（音の大きさ、リーリースタイムなど）の抽出を行う。信頼できる部分に関しては、一意的に対応を決定してゆく。信頼できない部分があると、その部分に関して動的計画法により探索を行なう。音の大きさ情報に関してはその音の時間周波数マップ上で音の大きさを読み取り、ルックアップテーブルを参照してMIDI規格の音の大きさ情報に変換する。和音内の音の微妙な時間的なずれ、音の持続時間に関しては、ノードにおける探索が終わってから再度探索を行なう。

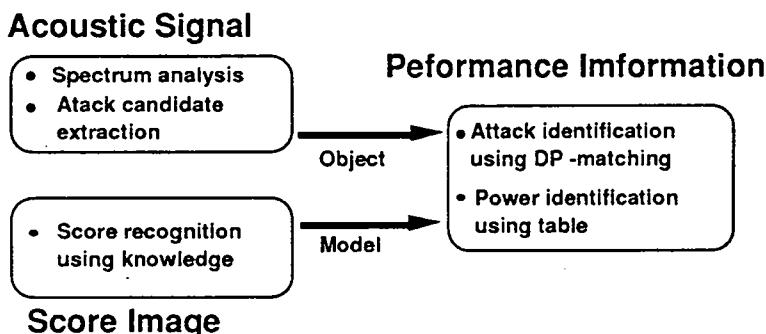


図5-18 演奏情報の同定

時間周波数マップは採譜処理で述べたのと同様の手法で得られる。時間周波数マップ上のノードを見つけるにはまず音の立ち上がり候補を見つける必要がある。ここでは音の全パワーに対してキャニーのエッジオペレータを用いてエッジ検出を行い、一段目の探索の音の立ち上がり候補のノードとする[Canny 1986]。また、各音階のエンベロープに対しても同様の処理を行い、1段目の対応探索のための尺度、そしてその時間軸に対するプロダクションを動的計画法における探索のノードとして用いる。楽譜情報はシンボルデータであるため音の立ち上がり時刻をそのままノードとして用いることができる（図5-19）。

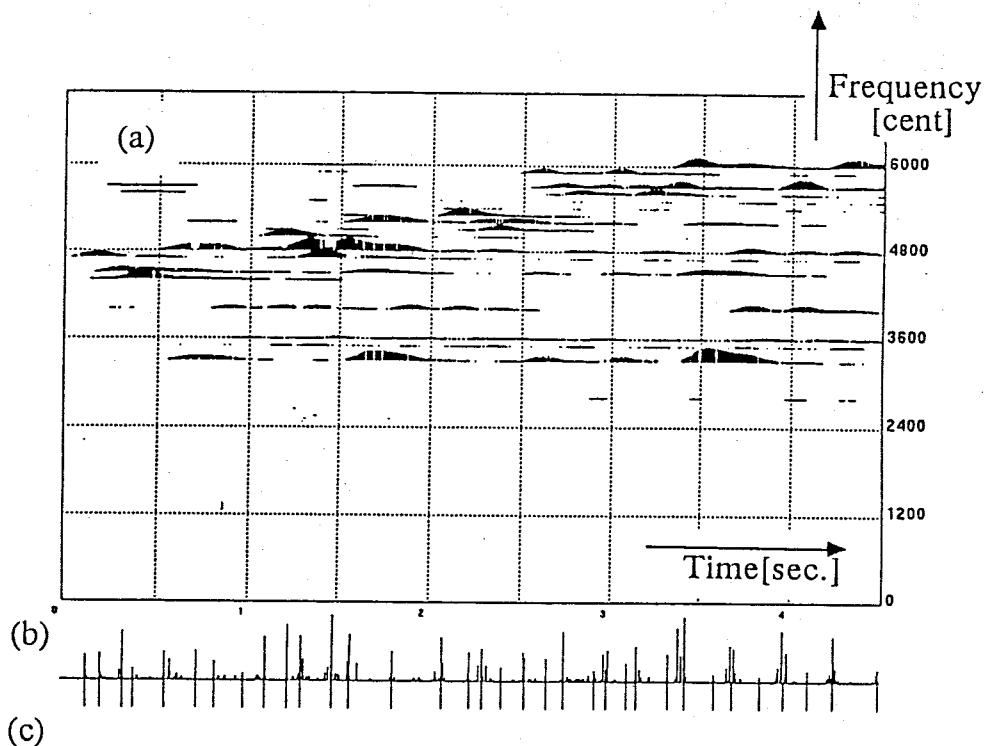


図5-19 ノードの選定

a)時間周波数マップ b)すべての周波数成分に対し検出した音の立ち上がり候補 c)全パワーに対して検出した音の立ち上がり候補、一段目の探索の音の立ち上がり候補として使われる

5-6-2 一段目の対応探索

基本的には、立ち上がり時の時間周波数マップ上の周波数構造 $P[i]$ と楽譜データから類推される周波数構造 $S[i]$ の類似性を用いて対応探索を行なう。ここで、 $P[i]$, $S[i]$ は半音階 i (C_0 を0, C_3 を36) を変数とし、その周波数(音階)での立ち上がりがある時は1、立ち上がりがない時は0を値としてもつ配列である。

以下にアルゴリズムを示す。

Step 1 現在、 n 番目のノードに関してマッチングがとれているとする。

Step 2 楽譜情報を用いて、時間周波数マップ上で $n+1$ 番目の音が始まると予想される範囲で明確な立ち上がりを探す。明確な音の立ち上がりは音の全パワーに対してエッジとして求めたもののうち、全エッジの平均強度以上もので与えている。

Step 3 時間周波数マップ上で明確な立ち上がりが発見できれば、その近隣時刻(最小

拍の半分以下)に明確な立ち上がり候補が存在する音高を抽出し, $P_n[i]$ を作成する。

Step 4 楽譜データから, そのノードにおける周波数構造(ここでは第3高調波まで)を推論し, $S_n[i]$ とする。

Step 5 $S_n[i] = P_n[i] = 1$ となる個数から $S_n[i] \neq M_t[i]$ となる個数を引いたものが正となるときに対応がとれたする。対応が取れたときは n に $n+1$ を代入し, Step 2 に戻る。対応が取れなかつたときには, $n+2, n+3 \dots$ 番目の音の対応を探索してゆく。

5-6-3 二段目の対応探索

上記のアルゴリズムにおいてもれなく対応が決定していく条件は, 明確な対上がりが発見できることと Step 5において評価基準を満足することである。実際の対象にはノイズ・音の揺らぎなどがあり, これらに対処するためにしきい値をあげる必要がある。すると今度は柔らかく演奏される部分の対応がとれなくなってしまうことがあるため, その対策として, 2段目の対応探索として動的計画法を用いた対応探索(DPマッチング)を行なっている。動的計画法とは連続性と単調増加性を満たす時系列パターン格子グラフから最適パスを効率的に求める手法として, 音声認識に積極的に用いられてきた方法である。音楽も単調増加性が保証されている時系列信号であるため, 探索法として動的計画法を用いるのは有効である。

1) DPマッチングの尺度

本手法はシンボルと信号のマッチングであるために通常のDPマッチングに用いられる差分法を適用することは難しい。ここでは, 各音のパワーエンベロープがもつ面積, すなわち時間周波数マップにおいてモデルと一致する音の大きさの時間積分値をマッチングの尺度として用いる。すなわち, この尺度が大きいほど, 整合度がとれているということになる。具体的には, 探索平面上の2つのノード (i, j) と (k, l) において, 尺度(整合度) $d(i, j, k, l)$ はモデルノード i がもっている音の数を n として

$$d(i, j, k, l) = \frac{1}{n \cdot p(i, j, k, l)} \sum_{m=1}^N \sum_{t=0}^{T_d(i)} \text{power}_f(t_o(m) + t)$$

$$p(i, j, k, l) = \frac{\max[t_m(k) - t_m(i), t_o(l) - t_o(j)]}{\min[t_m(k) - t_m(i), t_o(l) - t_o(j)]}$$

で与える。ここで $\text{power}_f(t)$ は, 各ピッチ f ごとに求めたパワー時系列である。 $t_m(i), t_o(j)$ はそれぞれモデルノード i , オブジェクトノード j における時刻を表し, $T_d(i)$ はモデルから得られる拍数に対応した時間を表している。 $p(i, j, k, l)$ は, モデル-オブジェクト間の時

間ひずみに対するペナルティである。

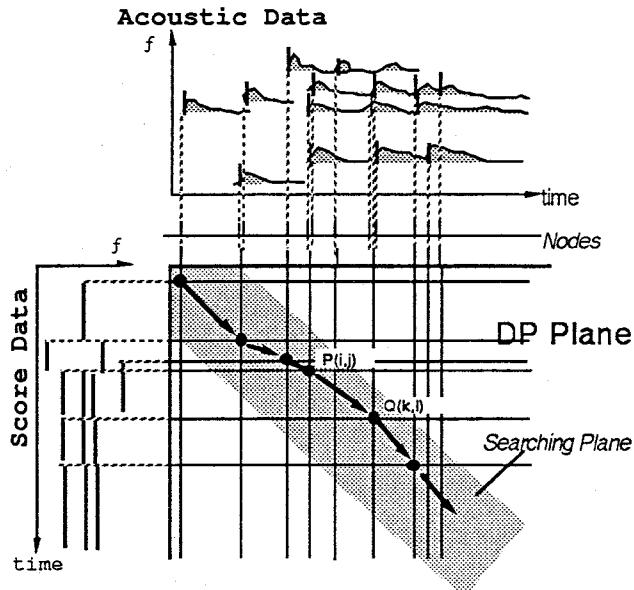


図5-20 動的計画法

2) 動的計画法のアルゴリズム

図5-20においてNode (k, l) からNode (i, j) までの類似度を $d(i, j, k, l)$, 始点からNode (i, j) までの整合度の最大値を $D(i, j)$ とする。DPのアルゴリズムは

$$D(0, 0) = 0$$

$$D(k, l) = \max[d(i, j, k, l) + D(i, j)]$$

$$k-m \leq j < k, l-n \leq i < l$$

で、順次、整合度が $D(k, l)$ 最大となるものを選んで行くという過程で与えられる。ここで m, n はそれぞれ楽譜ノード、実際の演奏のノードにおける読み飛ばしを許す個数であるが、モデルが絶対に正しいという仮定から $m=1$ である。

また、実際の演奏においては、拍の長さの変動はさほど大きくないため、実際の探索空間は図5-20における斜線部のように限定され、一層効率的に対応決定が行われる。

図5-21にモーツアルト作曲：トルコ行進曲の前半8小節分のマッチング結果を示す。

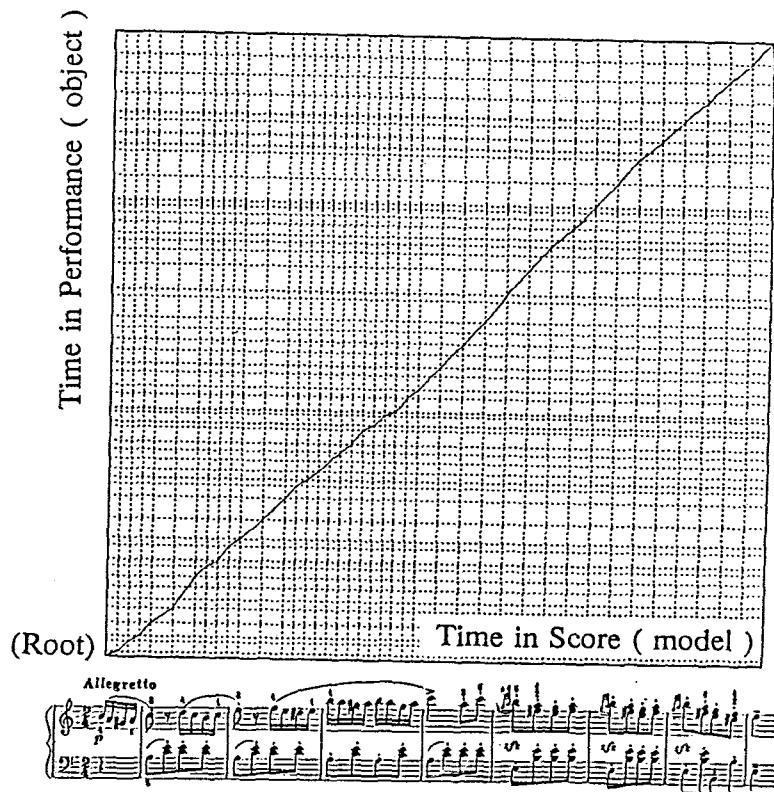


図5-21 マッチング結果（探索平面）

5-6-4 和音内の音の立ち上がり時刻、音の長さ・パワーの同定

単音の場合は対応の取れた点での時刻をそのまま立ち上がり時刻として採用する。和音に関しては、最も強度の強い立ち上がり候補の時刻がノードの代表として採用されるためにミクロな立ち上がり時刻を検出したことにはならない。和音については各音高について、その時刻付近で最もエッジ強度の強い点をそれぞれの立ち上がり時刻とする。

パワーについてはその音の立ち上がり直後の最大のパワーを検出し、予め用意したテーブルを用いてMIDI規格のベロシティに変換する。音の終了時刻は経験的に上記のパワーの1/30になった時刻で与えている。

5-7 表現ルールの学習

表現ルールは基本的には同じ音楽プリミティブに対して複数の共通した演奏法が発見されたときにルールとして学習される。以下、それぞれの音楽プリミティブの表現法の学習について述べる。

5-7-1 表意記号の演奏法の学習

楽譜中に記述されるクレシェンド、スタカートなどの記号の演奏法は、どのような機能を持つものかが予めわかっているために、その記号が付けられた部分の平均データからパラメータを同定するという形で比較的簡単に学習される。

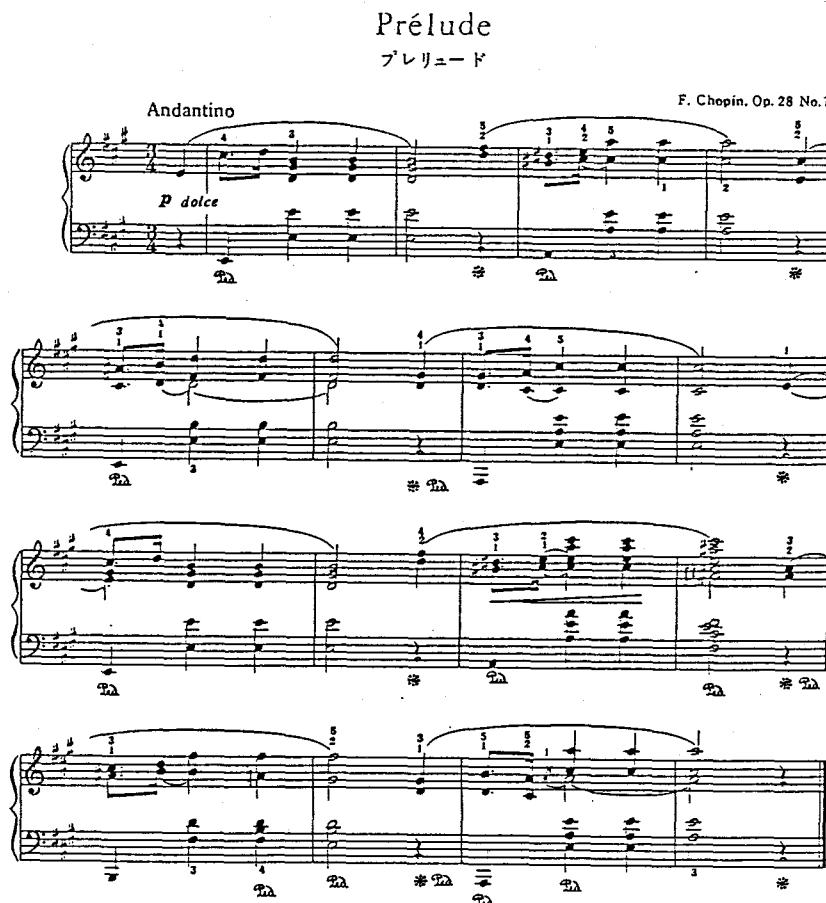
例えば、スタカート・テヌートに対しては、それに対する音長の長さの平均値を同定するということでルール化が行われる。しかし、音価や基本的なテンポに応じてその働きは変わるので、これらを変数としてルール化を行なう必要がある。各音価に応じた表現法として抽出したパラメータを以下に示す。

表7-1 音長のタイムテーブル

	音価	標準	スタカート	テヌート
全音符	4. 0	3. 85		
符点2部音符	3. 0	2. 92		
2分音符	2. 0	1. 92		
4分音符	1. 0	0. 92	0. 25	0. 98
8分音符	0. 5	0. 44	0. 125	0. 48
16分音符	0. 25	0. 19	0. 008	0. 23

5-7-2 モチーフの演奏法の学習

モチーフについては、それぞれがどのように演奏されるかは前もってわかるものではないので、フォームの各音のキー・オン、キー・オフ、ペロシティの相対的な変化の付け方が直接的に計算される。そして演奏ルールは同じタグのついたグループにおいて共通の表現（変化のある部分）が複数回検出されたときに生成される。この様子を図5-2-2に示す例を使って説明する。この曲はシンプルであるために、前節で述べた方法を用いることにより、矛盾のないグループが抽出されている。この楽曲の実際の演奏例からテンポ情報を抽出したものを図5-2-3に示す。



a)

Tag no.	Repetition beat	Groups (starting note, ending note)
1	6	(0, 5) (6, 11) (12, 17) (18, 23) (23, 29) (30, 36) . . .
2	12	(0, 11) (12, 23) (24, 35) -
3	24	(0, 23) (24, 47)

b)

図5-22 ショパン作曲：プレリュード, Op.28, No.7 a) 楽譜, b) グルーピング

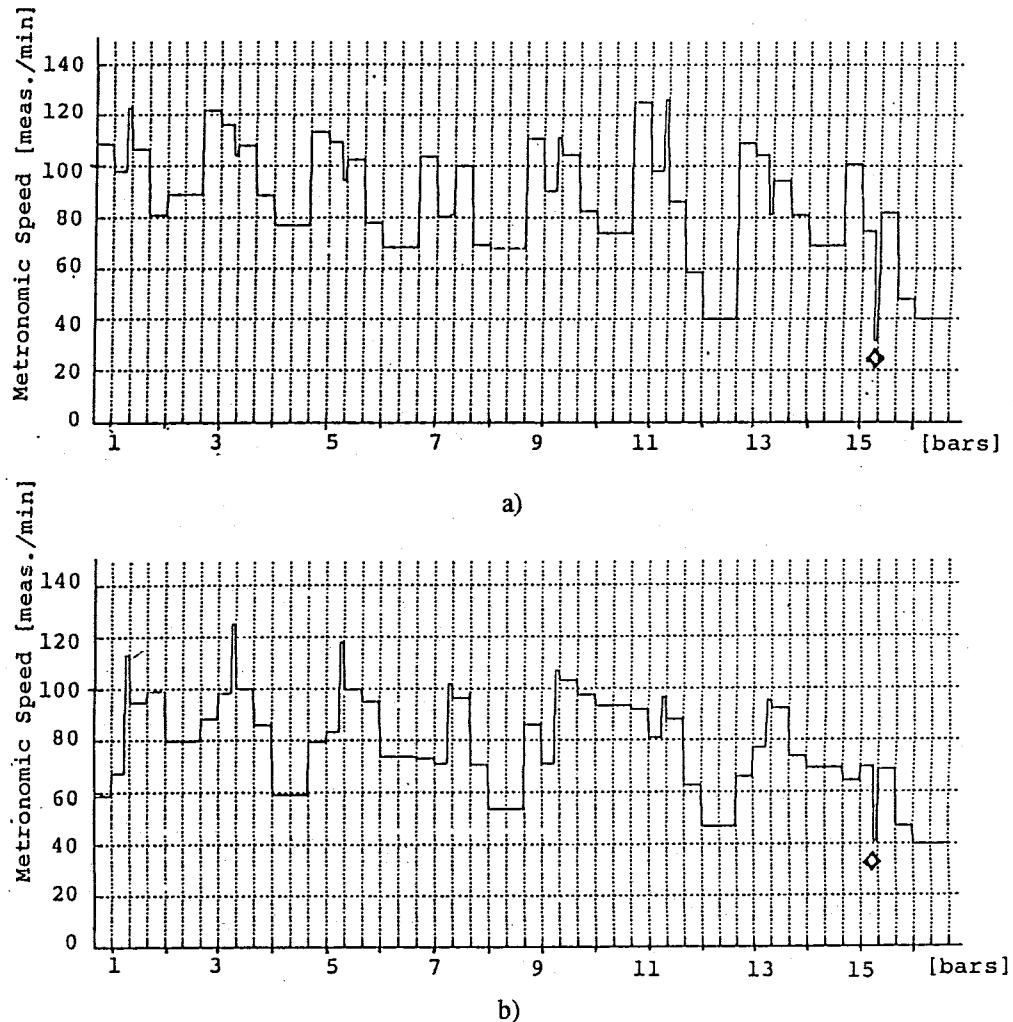


図5-23 演奏における速度分布 a) Pollini, b) Argerich

これより、この楽曲は、のところで顕著な特徴を持っていることがわかる。の表記上の長さは3:1であるのに対し、演奏例a), b)の演奏上での $\frac{\text{quarter note}}{\text{eighth note}}$ の長さの比の平均はそれぞれ3.14:1, 4.19:1である。

同じタグの付けられたグループで、同じ演奏特徴を持つものに対してモチーフ表現の抽出を行うので、例えば、演奏例b)では、6個のグループに共通の弾き方があるTag1をもった音列の表現法が抽出される。次に、そのモチーフのなかでの特徴部の演奏、すなわち $\frac{\text{quarter note}}{\text{eighth note}}$ の弾き方が一般化され、すべての $\frac{\text{quarter note}}{\text{eighth note}}$ のデフォルト表現として登録される。もし、他の学習段階で、例外が見つかったときには、それを含むより大きなモチーフ、

フレーズの形がその弾き方を規定しているすることにより整合性を保持して行く。例えば、予め  の弾き方が 2.7 : 1 であるというルールが既にあるときに図 5-23 に示すような例が提示されたとすると Tag1 の形、すなわち、2 小節分全体が  の弾き方として規定するものとして新たにルールが登録されて行く。

このほかに階層的な表現ルールを抽出するためのヒューリスティックとして、次に示す戦略を用いている。

1. 演奏特徴ルールを見つけだすときには小さなグループから大きいグループへと注目してゆく。
2. モチーフの繰り返しには特に注目し、音の大きさ、テンポのモーメントの変化のルールを抽出してゆく。(優先度は 1 の方が高い)

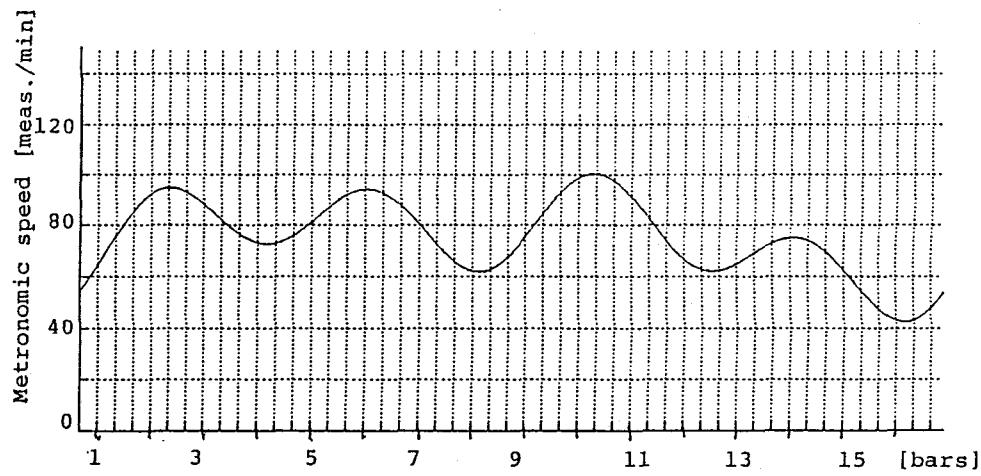
1 については、見つけた特徴を相殺してゆき、残った特徴に対し、再び、表現ルールを求めてゆくという手続きでルールの学習を行なってゆく。

ここでは、テンポ情報について主に説明を行なってきたが同様のことは音の大きさ情報についても行われる。

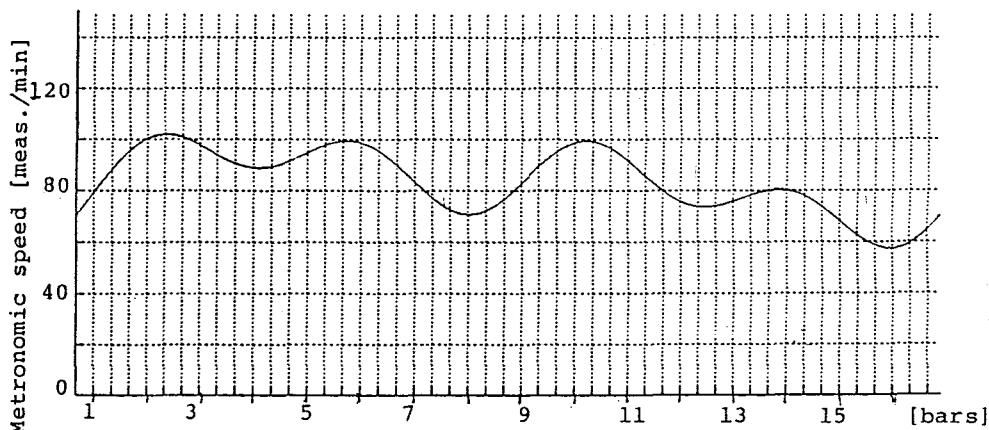
5-7-3 フレーズの演奏法の学習

フレーズの表現に関しては、モチーフの表現法を除去した演奏データに対し、ベータ関数をフィッティングさせるという方法でパラメータを抽出する。4 小節、8 小節など 2 の n 乗となるフレーズでのそれぞれの速度分はメトロノーム速度分布を一次元の波形と考え、フーリエ変換を行ない、それぞれの長さに応じた成分を逆変換することで分離が行なわれる。この波形に対し、先ず、振幅から G を推定し、ピークの位置から p1, p2 を推定してやればよい。ただし、2 小節以下の成分にはモチーフ等の影響もあるので、4 小節、8 小節、16 小節などの成分を除いたすべての、すなわち、フーリエ変換の高次の奇次成分をもって、2 小節の表現としている。

上記のプレリュードの演奏に対し、フーリエ変換後の第 0, 1, 2, 4 成分と第 0, 8, 16, 24, 48 成分に逆フーリエ変換を行なって生成したデータをそれぞれ図 5-24, 図 5-25 に示す。前者は 4 小節以上のフレーズの表現、後者は 2 小節以下の表現を示したものとなっている。



(a)



(b)

図5-24 4小節以上の表現) Pollini, b) Argerich

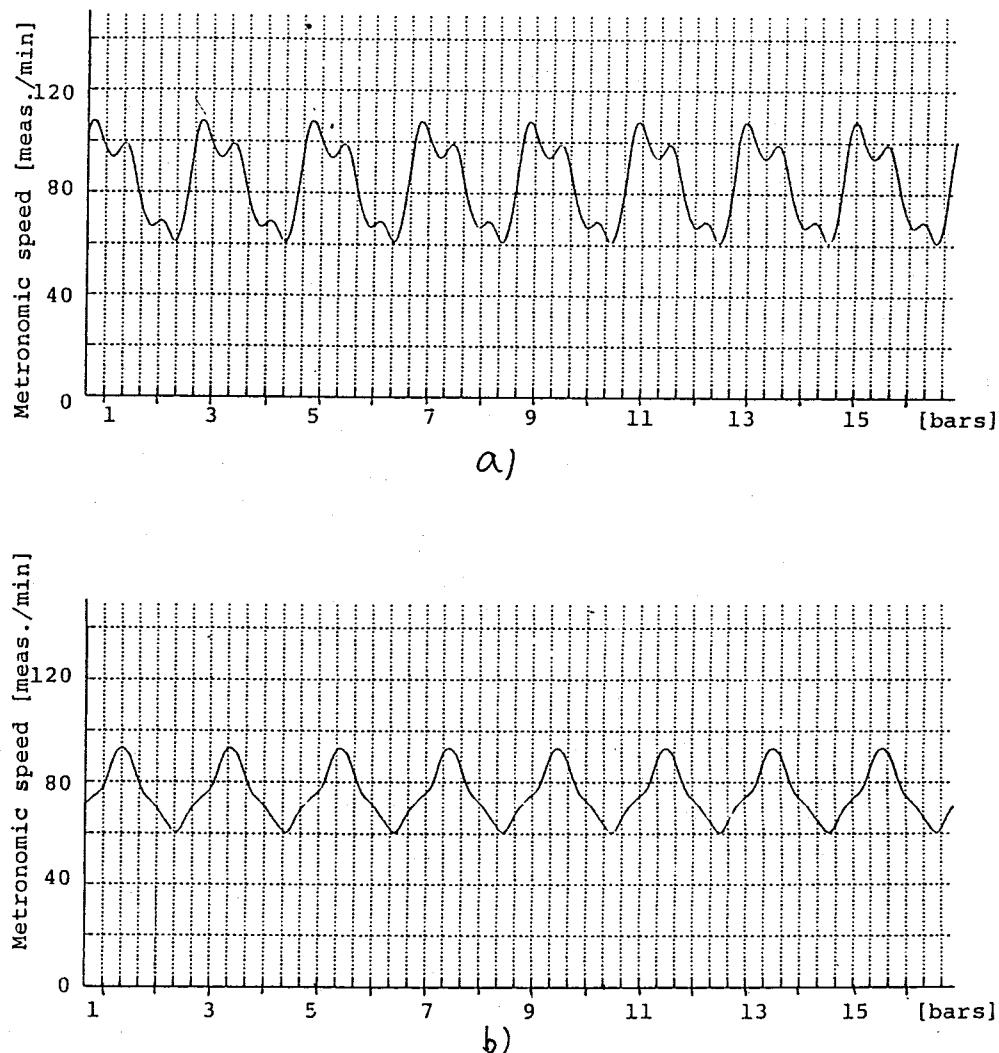


図5-25 2小節以下の表現 a) Pollini, b) Argerich

図5-24に示すデータ、すなわち、4小節以上の表現に関しては二人の演奏ほとんど同じ傾向が見られるが、図5-25にあげた2小節以下の表現に大きな差が見られ、非常に興味深い結果となっている。これは、フレージングに対しては、一般的な演奏法があり、むしろ、モチーフの表現については演奏者の個性が見られるというデータの一つとなるものである。

5-8 グルーピング・ストラテジーの抽出

各小節の最初の音は強拍といわれ、強く弾かれるることはよく知られている。これを実際の演奏に基づいてもう少し一般的にいって、グループをなす音列の先頭の音は何らかの形でそれとわかるように表現されるということになる。

人は、いろんな演奏に接して自分の表現法を身に付けてゆくが、単にグループとしての音列をどのように表現するかだけではなく、グルーピングの仕方も演奏例から学んでいると考えられる。その、手掛かりになるのが、グループをグループらしく見せる表現である。5-4節構造解析で示した例について考えてみる（図5-26）。

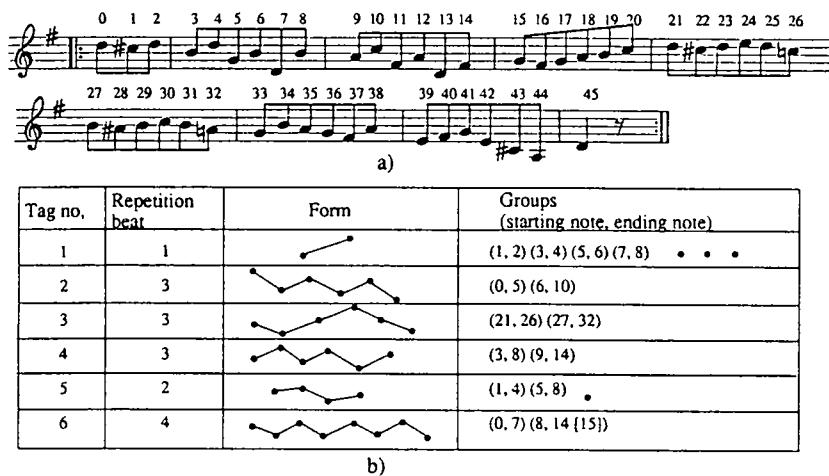


図5-26 多重のグルーピング

一般的な演奏においては拍子の1拍めにアクセントが付けられる。この例においては、3, 9, 15番目の音に対してアクセントが付けられるのが通常であり、これはこの曲が3拍子系の曲であること、すなわち、Tag.3を支持するものである。が、実際には、0と8の音に対してアクセントが与えられる演奏例もある。この場合、小節を単位として表現ルールの学習させようとする事は不可能である。0と8の音にアクセントがあるのは、演奏者が(0, 7), (8, 15)というグループに対してまとまりを感じ、それを表現しようとしているからである。MISでは、拍子をこえたグルーピングの候補の抽出を行っているので、この場合、Tag.8の形のものが演奏家によって選択されたと解釈される。学習段階で、このような例が与えられると、Tag.8の形のモチーフがあれば優先的に採用するというようなグルーピングストラテジーが生成されてゆく。もちろ

ん全ての楽曲においてこれが成り立つというわけではないが、この機構により解決できる事項は多い。また、学習のためのスケジュールを人間が考慮するという前提があれば、効率的に学習が進めることができる。

5-9 演奏ルールの抽出に関する検討

以上、この章の後半にかけては、演奏ルールの抽出について述べてきた。ここで述べた方法により、全ての演奏法が学習されるわけではない。しかし、少なくとも自律的に演奏ルールを学習する機構を持った自動演奏システムは従来になかったものであり、ルールベース型の演奏生成システムにおけるルール抽出という機構としては満足の行くものである。今後は学習対象例を増やし、より、一般的なシステムに向けて研究を進めて行く必要がある。

問題点としては、システムに提示する演奏例の順番を考慮してやる必要があることがある。もし演奏例の順番を考慮しなければルールの学習の効率が悪くなったり、最悪の場合間違ったルールを学習してしまうことがある。しかし、人間に演奏法を教えるときにおいてもスケジュールを考慮することが重要であることから、この仮定はさほど厳しいものとはいえないだろう。今後の課題としては、蓄積したルールベース群から与えられた条件を満たす最小ルールセットを探索する方法の実現が望まれる。また、音楽の構造に関しては現段階では、メロディに対してのみ行なっているため、伴奏を含めた総合的な解析機構を実現することが望まれる。

演奏者によって解釈は当然異なるし、同じ演奏者でも同じ構造をもった曲、さらには同一の楽曲でも全く異なったように演奏されることがある。本来、音楽の芸術としての興味からはこの部分を無視して考え得るものではなく、最終的な課題としてはこのようなことまでも含めた考察が必要である。しかし、特殊な部分を知るために一般的な部分を押さえておくことが不可欠である。現段階では不十分な点もあるがその一歩として、本研究を位置づけることが出来る。

5-10 まとめ

本章で述べてきた音楽解釈システムのまとめを以下に示す。

1) 音楽解釈MISの概要

- ・楽譜を入力として、情緒溢れる演奏を行なう。
- ・演奏法を自立的に学習する。

2) 演奏データの生成

- ・楽譜認識
- ・構造の解析
- ・ルールによる演奏データの生成

3) 構造解析

- ・拍子を越えたまとまりの解析
- ・表現を行うためのフレーズを選択するグルーピングストラテジー

4) 表現法の抽出

- ・楽譜情報を利用
- ・音の立ち上がりに注目した2段の対応探索

5) 演奏ルールの抽出

- ・パラメータ学習による表意記号表現の抽出
- ・モチーフ表現の抽出
- ・フレーズの表現の学習
- ・グルーピングストラテジーの抽出

第6章 作曲過程

自動作曲は計算機の出現とほぼ同じくして取り組まれた音楽情報処理の中でも歴史のあるテーマとして位置づけられる。自動作曲システムの出初めの頃は、音楽的制約や取りうる状態をルールとして記述し、そのルールと乱数で音楽を生成するというものであったが、徐々に、処理に人間的思考のシミュレートをめざしたもの、あるいは、思考処理については完全に人間にまかせ音楽制作環境としての機能を追及したものなど、試みの段階を終え、質を追及する段階にはいっている。本章では、作曲・編曲過程における意志決定の一手法として、感性パラメータを利用した方法を紹介する。

6-1 作曲と創作について

第2章でも述べたように、自動作曲は、コンピュータの歴史の中でも非常に初期から試みられている。「イリアック」組曲以来、非常にたくさんの人達によって（個人レベルでの趣味の人も含めて）自動作曲プログラムが制作されてきたが、ほとんどが、コンピュータ上に何らかの音楽的秩序を与えるフィルタ（ルール）を構築して、ランダム信号、あるいは、モチーフを与えて音楽を作成するというものであった。このようにして作られる音楽の限界によるものか、アメリカやヨーロッパでは、作・編曲の支援システムという目的を全面に出した研究が進んでいる。その背景としては、哲学的には「創造は人間の楽しみなんだ。」ということと、技術的には感性的意志決定機構の問題がその根底にあるようと思われる。

さて、自動作曲の可能性だが、その限界は作曲システムというより聴取モデル（できた音楽を判断するモデル）が与えるものだと、筆者は考えている。作曲家に接する度に、「作曲は創、オリジナリティが最も重要な要素であり、従来からの音楽のわく（その人らしさでもよい）からほんの少し逸脱したところに面白い音楽がある」という意見を聞くことが多いが、これは、聴取者も納得できる意見であろう。この意見において重要なことは、作曲者が今まで曲の原形を作るための知識を越えた範囲でそれを評価しうる“耳”をもっているということである。このように考えるともし、コンピュータでランダム信号をふんだんに使って偶然にできてしまった曲でも、その作曲家が「これだ」と思えば、十分に立派な作曲過程であるという帰結を得ることができる。

コンピュータは与えた音楽秩序の世界での（芸術的には可でも不可でも）作品候補を作ることはできる。また、もう少し創作的にするなら、与えた音楽秩序を少々逸脱した作品候補もエグゾースティブ・メソッド的に数え上げることができる。しかしながら、人間的に働くためにはでき上がった作品候補から、「良いもの」を選びだし、新しい音楽秩序を一般化して、次の作曲時に生かす機構が必要となる。現在のところ、このような機構をもった研究例は、存在していない。非常に難しいテーマではあるが、認知科学、AIなどの複合領域から見て、非常に興味深いテーマと言える。

6-2 作曲支援システムにおける感性情報の利用

前節では、作曲や創作について、筆者の考え方について紹介した。今までに作曲システムや編曲システムとして紹介された研究はかなりの数にもはあるが、聴取モデルの存在を創作の必要条件とすると、創作を行なう作曲プログラムは存在していないことになる。自立的に意志決定を行なうプログラムであっても、作成された作品群を評価し、選び出すということは出来るものではない。その作業 자체を人間が行なっていることを考えると、現存する作曲あるいは編曲システムはすべて支援システムということになろう。

それでは、作・編曲支援システムは創作システムと比べると、劣ったものであるかというとそうではない。人間における音楽助手を考えて見るとよくわかる。例えば、ある作品におけるモチーフをすべて、置き換えるといふタスク一つ取っても、全く同じ音の並びしか考えない場合は、比較的簡単なマッチングを行なうことにより対処できる。しかし、ある程度変形を許して機能というレベルにまで範囲を広げると、それこそ、音楽に対する深い知識が要求される。このような機能を余すことなく実現することと幼児レベルの創作過程を実現することのどちらが難しいとは、そう簡単に決めれるものではない。作・編曲支援システムにおいても、システムが音楽をわかっていることは非常に重要なことである。以下、そのような作・編曲システムの一つの形として、感性パラメータを利用した音楽エディティングシステムを紹介する（図6-1）。

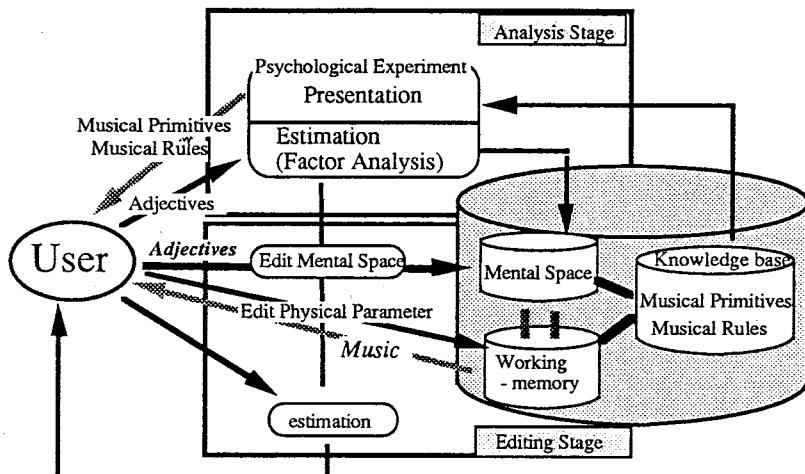


図6-1 音楽エディティングシステム

このシステムにおけるメンタルスペースは第4章で述べたファクターモデルに相当するものである。ユーザはファクターモデルの座標すなわち感性パラメータによって、ワーキングメモリに記憶されている音楽プリミティブ、および、音楽データの演奏例を検索することができる。

6-2-1 作曲支援

作曲支援については様々なアプローチが考えられるが、コード進行で曲の枠組みを規定する作曲形態の支援を以下に示す。

- 1) このような雰囲気の曲を作りたいというイメージをメンタルスペースの座標として、システムに提示することにより、そのイメージに（距離的に）近いコード進行がいくつか提示される。ユーザがコード進行を選択する。
- 2) 次に、基本的なリズムパターンを選択肢から、ユーザが適当なものを選択する。リズムパターンについては、マーチ風、ニューミュジック風、演歌風などというよう選択肢自体が少ないので、そのままの呼び名でも十分であると判断している。
- 3) コード進行とテンションに関する情報と基本的なリズムパターンが決まれば、メロディを形づける構造音が限定される。この構造音の選択肢から、ユーザーが適当なものを選び出していくことによって、メロディの骨格が出来上がる。

- 4) 骨格情報から、適当な音楽ルールで経過音を埋めていくことで、メロディ、リズム、コード進行からなる音楽を生成される。この時に、第4章で述べた予期モデルを使うことによって、その予測性、すなわち、期待感という感性パラメータを利用するとの可能である。
- 5) 出来上がった音楽をユーザーが評価したうえで、直したい部分があれば、エディティングを行なう。なお、ユーザーが生成されたメロディのについては、予期モデルによって評価することも出来る。

6-2-2 編曲支援

前節では作曲支援の例を取り上げたが、編曲支援においてはもっとシンプルにデータの検索、探索空間の限定という目的で感性情報が利用される。例えば、図6-2に示す音列にコードを付けること（和声付け）を考えてみよう。



図6-2 和声付けの問題

通常、和声付けを行なうときの最も基本的なルールは、強拍における音符に調和したコードを対象曲の調性に適合したコード群から選ぶというものである。図6-2の場合だと、1小節目、2小節目に適合するコード進行として、 $\{G, G\}$ という候補が最も単純なものとして選ばれるだろう。その他に、 $\{G, C\}$ ， $\{Em, G\}$ ， $\{Em, C\}$ なども考えられるが、テンションを含んだコードも取りうるとすると、非常に大きな探索空間を持つことになる。このような時に、音楽の経験に乏しい人にとって、例えば、 $\{CM7, Gaug\}$ などのコードを、自分のイメージと対応させて直接的に選ぶということは不可能である。ところが、予め、取りうるコードのイメージ空間をもってシステムと交信できれば、イメージにあったものを一回で選びだせないとしても、探索空間をかなり縮小することが出来、聴取の上で、気に入ったものを選ぶことが可能となる。ここで問題となるのは、シンタックスである。音楽プリミティブはその組み合わせ方

によってイメージが異なってしまうという問題であるが、人間がそう感じる以上、どんな方法を取っても、システムが推測できるものではない。これに関しては、出来るだけ多くの音楽プリミティブ、およびその組み合わせ方に関して、そのイメージを登録して行くより方法はない。しかし、探索自体は、イメージを手掛かりにするため、データが増えたからといって、極端に、探索能率が落ちるわけではない。

6-2-3 メディアの変換

以上の例は、音楽制作の支援という目的でミュージックエディティングシステムを利用する場合であるが、入出力データとして、感性情報を利用していることから、この感性情報を媒体とすることにより、言語、画像など他のメディアとのリンクが可能となる。一つの応用としては、音楽療法用のプログラムの自動生成などが期待される。音楽療法は音楽聴取を行なうことによって精神状態の改善を行なう、心理学的療法の一種である。音楽療法を行なうときには、専門家が、患者をよりよい精神状態に近付けるように精神状態の推移や方針の処方せんを作成する。そして、その処方せんに基づいて音楽の種類と順序を決定するという作業が取られるが、処方せんの段階で、感性情報がパラメータ化された形で記述されれば、最も適した音楽プログラムを自動的に作成することも可能である。ほかに、ある音楽にあった画像の選択、あるいは、その逆に、画像にあった音楽の選択など、マルチメディアにおけるデザイン支援システムとしての可能性がある。

6-3 まとめ

本章では、感性情報処理の作・編曲支援システムへの応用について述べた。

- 1) 作曲と創作
- 2) 感性情報を利用した音楽エディティングシステム
 - ・作曲支援システム
 - ・編曲支援システム
 - ・メディアの変換における応用

第7章 結論

本論文は音楽情報処理、感性情報処理という二つの情報処理分野のもとに、より自然な形の知的音楽助手としての機能をもつ音楽システムの実現に向けて、著者が大阪大学基礎工学部大学院修士課程、博士課程を通して研究してきた成果をまとめたものである。

第1章では、本研究の指針を明らかにするために、その背景となる音楽情報処理、感性情報処理について述べた。音楽情報処理に関しては、今まで幅広く取り組まれてきた領域での研究例を示しながら一般的な概観を行なった。感性情報処理については、感性を情報として扱っていくための枠組みを与え、感性情報処理の対象、その意義について述べた。

第2章では、まず、従来の音楽情報処理において感性的側面があまり考慮されていなかったことについて指摘した。その上で、音楽感性情報処理の目的が音楽を理解し、人間に見られる信号の入出力機能を実現することであるということを述べた。そして、本研究において構築を進めてきた音楽感性情報処理システム“KANSEI”の概要について述べた。

第3章では、“KANSEI”的耳に相当するものとして、単一楽器の多重音の採譜システム、および、複数音源の分離機構について述べた。前者はピアノの音響入力に対しほぼ90%以上の認識率を上げており、システムにおけるデータ入力の重要な機能となっている。後者については、モデルとタスクによって自動的に最適な処理が実現されるという他に例を見ない音響認識機構の提案を行なうものである。実験例としては、ピアノ音とクラリネット音の分離、楽器の判定について示している。これらは、単なる自動採譜システムというだけではなく、音楽理解モデルの楽音知覚機能としての意味合いをもつものである。

第4章では、“KANSEI”的心の部分に相当する部分について述べた。ここでは、まず、採譜処理により抽出されたノートデータからどのような機構でコード進行、調性、リズムパターン、メロディなどの音楽プリミティブが認識されるかを示した。そして、音楽プリミティブと形容詞情報とのリンクによって、感想文を出力する処理を紹介した。次に、より、人間に即した機械のための音楽聴取モデルを示した。この聴取モデルにおける感性状態の記述法として、ネットワーク型に形容詞および音楽プリミティブをリンクした伝播モデルと因子分析法により形容詞空間を小数のパラメータで記述するファクタモデルについて述べた。さらに、能動的な聴取過程に位置づけられ、「興」にあたる

感性情報を抽出する予期モデルについて述べた。予期モデルの実現に関しては、できるだけヒューリスティックな部分を軽減して、情報量を圧縮するという原理にしたがって構造を認識する手法を提案している。

第5章では、音楽解釈機構の実現に焦点を当て、楽譜入力からの情緒溢れる演奏データの生成過程および演奏ルールの学習過程について述べた。情緒溢れる演奏を行なうことを目指したエキスパートシステムは2, 3紹介されているが、演奏ルールを自立的に、しかも、実際の演奏例から学習するという機能を有したものは現在唯一といってよい状態である。この機能は一種のセンスの学習に位置づけられるものであり、感性情報処理の最も興味ある課題の先駆となるものであろう。

第6章では、作・編曲過程について述べた。章の冒頭では作曲というテーマを考えていくに当たり、音楽創作に対する著者なりの定義を示した。次に、感性情報を利用した音楽エディティングシステムを紹介し、その作曲支援システム、編曲支援システムとしての応用、さらに、マルチメディア処理における応用について述べた。この章は、簡単ではあるが、感性情報処理システムにおけるインターフェースの実例を示したものとなっている。

感性情報処理はまさに黎明期に相当する領域であり、何を研究すべきか、どのように研究していくべきかについては現在明確な定義はない。本論文では、処理的実現性という観点から、感性情報をパターン感性情報、シンボル感性情報、パラメータ感性情報の3つに分類し、それぞれの感性情報の抽出、生成、変換過程を示すというアプローチによる取り組み方を示した。論文中では、採譜処理、演奏情報の抽出などがパターン感性情報の抽出に相当する。そして、音楽解釈システムにおける演奏の生成がパターン感性情報の生成に相当する。シンボル感性情報の扱い、パラメータ感性情報のハンドリングについては、第4章の音楽分析と理解として取り上げている。また、第6章作曲過程においては、感性情報の具体的な利用法を示したものとなっている。

音楽が学問として取り上げられ始めたのは、古代、哲学が生まれた時期とほぼ一致する。それ以来、人間、芸術、数学の接点で音楽に対する弛まぬ考究が続けられてきた。その歴史の上で近代音楽を扱っている学問的領域は、音楽学はもちろんのこと、美学、哲学、心理学などの分野がある。これに対し、工学的な手段で音楽が論じられるようになったのはつい最近のことである。工学、なかでも、パターン認識を専門としてきた著者は音楽を研究対象を扱うものとして新参者であり、今まで音楽を対象に行なわれた研究すべてに目を通すことは不可能であった。本論文で扱ったモデル等に関して、他の歴史ある分野によってすでに解決されてたり、あるいは、さらなる検討がなされている

可能性については否定する確証をもちえない。しかし、工学的アプローチについてのみ言えば、本論文で述べてきたように総合的な視点から音楽感性情報処理に取り組んだ研究は他に例はなく、このような対象を工学的アプローチで扱って行く上での困難、および、解決して行くべき事項、そしてその解決に向けての糸口は、かなり明確に示せたものとして自負している。

本論文が、今後の音楽情報処理、感性情報処理の発展に多少なりとも寄与できれば幸いである。

謝　辞

本研究を行う機会を与えてくださいり、日頃、研究の方針、内容について、細やかなご助言、ご指導を賜りました井口征士教授に心から感謝いたします。また、国内はもとより、国際会議での発表機会を数回にわたり与えていただいたことは、ともに黎明期にあたる音楽情報処理、感性情報処理の動向に対する理解を深め、研究方針を定めていくうえで、得がたい経験となりました。合わせて感謝致します。辻三郎教授は、いち早く感性情報処理の必要性を説かれ、本研究を遂行していくにあたり、温かい目で、支援してくださいました。ここに、深謝をささげます。豊田順一教授には、懇切なるご討議を賜りました。ここに、厚くお礼申し上げます。中村敏枝助教授には、音楽心理学、SD法などのついてご指導頂きました。深く、感謝の意を表します。今井正和助手（現、辻研究室）には、修士課程2年の時に、本論文においても中心を占める採譜処理、および、その基礎に位置する信号処理について、細やかな指導をいただきました。佐藤宏介助手には、筆者が、学部4年生から修士課程1年にかけて、画像処理、パターン認識の考え方を、指導いただきました。この経験は、音楽を対象としていく上でも、非常に、参考となりました。今井正和助手、佐藤宏介助手に対して、心から感謝の意を表します。加藤博一助手とは、学部時代から、クラブ、研究室と同期として過ごしてきました。研究面においては、計算機環境の整備という点で、特にお世話になりました。また、音楽感性情報処理システム"KANSEI!"の重要なデータ入力機能である楽譜認識は加藤助手の手によるものです。あわせて、深謝いたします。

本論文は、井口研究室音楽グループの手助けなしに、成り立つものではありません。その意味で、本論文は、昭和62年から、平成2年にかけての井口研究室音楽グループ全体の成果として位置づけられるものです。ピアノにおける演奏情報の抽出と解析に関しては、高見啓史氏（現、ヤマハ）、尺八をはじめとした伝統音楽の支援ツールに関しては、今井重彰氏、解釈システムに関しては、福岡俊之氏、作・編曲支援システムに関しては才脇直樹氏、音源分離システムに関しては、長束哲郎氏との共同研究としての成果です。以上の皆様に、深く感謝いたします。また、沼田卓久助教授をはじめ、日頃様々な面でお世話になっている井口研究室の諸氏に感謝します。

音楽情報科学研究会においては、現在、発展しつつある音楽情報処理の分野における新鮮な発表に接し、また、自分の発表においても、貴重なご意見を伺うことが出来ました。音楽情報科学研究会の幹事および会員の諸氏に感謝の意を表します。

最後に、研究生活を支援してくれた家族に感謝します。

参考文献

- [Ames 1982] Ames, C. : "Crystals: Recursive Structures in Automated composition", Computer Music Journal, 6(3) (1982)
- [Ames 1983] Ames, C. : "Stylistic Automata in Gradient", Computer Music Journal, 7(4) (1983)
- [Andronico 1982] Andronico, A. and Ciampa, A. : "On Automatic Pattern Recognition System and Aquisition of Printed Music", Proc. ICMC, p.245 (1982)
- [Anzai 1987] 安西祐一郎：認知科学と人工知能，共立出版 (1987)
- [Byrd 1977] Byrd, D. : "An Integrated Computer Music Software Syestem", Computer Music Journal 1(2), 55 (1977)
- [Beyls 1988] Beyls, P. : "Introducing Oscar", Proc. ICMC, pp.218-230 (1988)
- [Blosh 1985] Bloch, J. and Dannenberg, R. : "Real-Time Computer Accompaniment of Keyboard Performances", Proc. ICMC, p.279 (1985)
- [Canny 1986] Canny, J. : "A Computational Approach to Edge Detection", IEEE Trans., Vol-PAMI, No.8 (1986)
- [Chadabe 1984] Chadabe, J. : "Interactive Composing", Computer Music Journal, 8(1) (1984)
- [Chafe 1986] Chafe, C. and Jaffe, D : "Source Separation and Note Identification in Polyphonic Music", ICASSP86, pp.1289-1292 (1986)
- [Chowning 1973] Chowning, J.M. : "The Synthesis of Complex Audio Spectra by Means of Frequency Moduration", Jounrnal of Audio Eng. Soc., pp.526-534 (1973)
- [Clynes 1984] Clynes, M. : "Secretes of Life in Music", Proc. ICMC, pp.225-232 (1984)
- [Choi 1990] チョイ，原島，武部："顔の3次元モデルに基づく表情の記述と合成"，信学論，Vol.J73-A, No, 7 pp.1270-1280 (1990)
- [Cointe 1983] Cointe, J.P. and Rodet, X. : "Forms:Composition and Scheduling of Processes", Computer Music Journal, 8(3), pp.32-50 (1983)
- [Cope 1987] Cope, D. : "Experiment im Musical Intelligence", Proc. ICMC, pp.174-181(1987)
- [Dannenberg 1989] Dannenberg, R. : "Real-Time Scheduling and Computer Accompaniment", Current Directions in Computer Music Research, MIT Press, pp.225-262 (1989)
- [Desain 1989] Desain, P. and Honing, H. : "The Quantization of Musical Time: A Connectionnist Approach", MIT press, Computer Music Journal, 13(3), pp.56-66 (1989)
- [Diener 1989] Diener, G. : "TTrees : A Tool for the Compositional Environment", Computer Music

- Journal, 13(2), pp.77-85 (1989)
- [Duetsch 1987] ダイアナ・ドイチュ：音楽の心理学，西村書店 (1987)
- [Ebcioğlu 1984] Ebcioğlu, K : "An Expert System for Schenkerian Synthesis of Chorales in the Style of J. B. Bach", Proc. ICMC, p.135 (1984)
- [Frydon 1984] Fryden, L. and Sundberg, J.: "Performance Rules for Melodies. Origin, Functions, Purposes", Proc. ICMC, pp. 221-225 (1984)
- [Fujii 1989] Fujii, H. et al : "A Music Arrangement System Based on Hummed Melodies", Proc. ICMPC, pp.103-106 (1989)
- [Fukuoka 1990] 福岡, 片寄, 井口："音楽解釈システム MISにおける演奏生成処理について", 情報処理学会前期全国大会予稿集 (1990)
- [Gabrielsson 1985] Gabrielsson, A. et al. : "Interplay between analysis and Synthesis in Studies of Music Performance and Music Experience", Music Perception, 3, pp.59-86 (1985)
- [Hara 1983] 原, 井口："複素スペクトルを用いた周波数同定", 計測自動制御学会論文集, Vol.19, No.9, pp718-723 (1983)
- [Harashinma 1989] 原島博："知的情報通信と次世代情報通信", コンピュータワールド'89, pp.127-134 (1989)
- [Hatanaka 1986] 畑中, 市川："音声認識におけるアルゴリズム", 電子通信情報学会誌, 69, 4, pp306-310 (1986)
- [Hewlett 1989] Hewlett, W.B. et al. : Computing in Musicology A Directory of Research, CCARH (1989)
- [Holtzman 1980] Holtzman, S. : "A generative Grammer Definitial Language for Music", Interface, 9 (1) (1980)
- [Hiller 1959] Hiller, L. et al.: Experimental Music, McGraw-Hill.
- [Hiller 1985] Hiller, L. and Ames, C. "Automated Composition: An Installation at the 1985 International Exposition in Tsukuba JAPAN", Persp. New Mus. 23(2) (1985)
- [Hiraga 1989] Hiraga, Y. : "A Computational Model of the Cognition of Melodic/Harmonic Progression", Proc. ICMPC, pp61-66 (1989)
- [Hirata 1987] 平田, 青柳："新世代音楽システム ICOTone", コンピュータと音楽, pp.101-104, BIT別冊 (1987)
- [Ikeuchi 1989] Ikeuchi, K. and Hong, K.S. : "Determining linear shape change: Toward automatic generation of object recognition program", Proc. Intl. Conf. of CVPR (1989)
- [International MIDI Association 1983] International MIDI Association : MIDI 1.0 Specification

(1983)

[Inokuchi 1980] 井口征士："音楽情報の処理—電算機を用いた自動採譜", 計測と制御, 19, 3, pp.314-319 (1980)

[Inokuchi 1988] 井口, 今井："音楽情報の認知", AIジャーナル, No. 5 (1988)

[Inokuchi 1990] 井口, 片寄："コンピュータと音楽", 電子通信学会誌 9月号, Vol73, No.9, pp.965-956 (1990)

[Iwashita 1979] 岩下豊彦：オスグッドの意味論と S D法, 川島書店 (1979)

[Kawahara 1988] 笠原裕："感性デザインシステム", Proc. Computer World '88, pp. 84-91 (1988)

[Katayose 1988a] Katayose, H. et al. : "Sentiment Extraction in Music", Proc. Intl. Conf. on Pattern Recognition, pp. 1083-1087 (1988)

[Katayose 1988b] 片寄, 今井, 井口："音楽における感性情報抽出の試み", 人工知能学会誌, Vol.3, No.6, pp.748-754 (1988)

[Katayose 1989a] Katayose, H. et al. : "An Approach to Artificial Music Expert, Proc. ICMC, pp. 139-146 (1989)

[Katayose 1989b] Katayose, H. et al. : "Music Interpleter in the KANSEI Music System", Proc. ICMC, pp.147-150 (1989)

[Katayose 1989 c] 片寄, 才脇, 井口："音楽鑑賞モデルにおける感性量の取り扱い", 人工知能学会全国大会, pp.543-456 (1989)

[Katayose 1990a] 片寄, 井口："知的採譜システム", 人工知能学会誌, Vol.5 No.1.

[Katayose 1990b] Katayose, H. et al.: "Expression Extraction in Virtuoso Music Performances", Proc. of Intl. Conf. on Pattern Recognition, pp. 780-784 (1990)

[Katayose 1990c] 片寄, 福岡, 井口："音楽解釈システムMISにおける演奏ルールの抽出について", 情報処理学会前期全国大会.

[Katayose 1990d] Katayose, H. and Inokuchi, S. : Kansei Music System, MIT Press, Computer Music Journal, Vol.13, No. 4, pp.72-77.

[Katayose 1991] Katayose, H. and Inokuchi, S. : "Music Recognition -Through Construction of Automaric Notation System-", Music Perception (invited to appear)

[Kato 1988] 加藤, 井口："小節単位処理に基づいたピアノ楽譜の自動認識", 電子情報通信学会論文誌, Vol.J71-D No.5 (1988)

[Kawahara 1981] 川原浩：実用鍵盤和声，音楽之友社 (1981)

[Kitagaki 1977] 北垣, 清水, 末武："コンピュータによる楽音認識の試み", 日本音響

会誌, 33-9 (1977)

[Kitahashi 1988] 北橋忠宏："工学における認識と理解", 電子情報通信学会誌, Vol. 71, No.11, pp.1113-1118 (1988)

[Kroland-Martinet 1987] Kroland-Martinet, R. et al.: "Analysis of Sound Patterns through Wavelet Transforms", Intl. Journal of Pattern Recognition and Artificial Intelligence pp.97-126 (1987)

[LaRue 1988] ラルー, 大宮:スタイルアナリシス,音楽の友社 (1988)

[Lerdahl 1983] Lerdahl, F. and Jackendoff, R. : A generative theory of tonal music, The MIT press, 368 pages (1983)

[Longuet-Higgins 1987] Longuet-Higgins, H.C. : Mental Processes, The MIT Press (1987)

[Martens 1985] W. Martens : "PALLETE : An Environment for Developing an Individual Set of Psychologically Scaled Numbers, Proc. of ICMC pp.355-365 (1985)

[Mase 1990] 間瀬 :"オプティカルフローを用いた表情の認識", 信学全大, D-544 (1990)

[Moorer 1975] Moorer, J.A. : "On the segmentation and analysis of continuous musical sound", Rep. STANM-3 (1975)

[Musha 1989] Musha, T. : "Music and Heartbeat Rate", Proc. ICMPC, pp.281-186 (1989)

[Mathews 1969] Mathews, M. : The Technology of Computer Music. MIT Press (1969)

[Mathews 1989] Mathews, M. : "The Conductor Program and Mechanical Baton", Current Directions in Computer Music Research, MIT Press, pp.263-282 (1989)

[Meyer 1956] Meyer, L.B. : Emotion and Meaning in Music, University of Chicago Press (1956)

[Minsky 1989] Minsky, M. : "Music, Mind and Meaning", The Music Machine, MIT Press, pp. 639-656 (1989)

[Mont-Reynaud 1989] Mont-Reynaud, B. M. and Mellinger, D. K. : "Source Separation by Frequency Co-Modulation", Proc. ICMPC (1989)

[Motita 1989] Morita, H. et al. : "Computer Music System Which Follows a Human Conductor", Proc. ICMC, pp.207-210 (1989)

[Murao 1987] Murao, T. : "楽曲分析における認知", 音楽と認知, 東京大学出版会 (1987)

[Murao 1989] Murao, T. : "Compositional Structure からの代償としてのCognitive Structure", 音楽情報科学研究会会報, NO.26 (1989)

[Nagamachi 1990] 長町三生:感性工学, 海分堂 (1990)

[Nakamura 1979] Nakamura, Y. and Inokuchi, S. : "Music Information Processing System and its Application to Comparative Musicology", 6th IJCAI (1979)

[Nakamura 1981] 中村勲 :"発音機構のシミュレーション(ピアノの音響楽的研究)", 日本音

響学会誌, 37, pp.65-75 (1981)

[Nakamura 1985] 中村敏枝："調の性格に関する音楽心理学的研究", 大阪大学教養部研究集録, No. 33 (1985)

[Narmour 1983] Narmour, E. : "Some major theoretical problems concerning the concept of hoerarchy in the analysis of tonal music", Music Perception, 1, No.2, pp.77-86 (1983)

[Niihara 1984] 新原, 今井, 井口："歌唱の自動採譜", 計測自動制御学会論文集, Vol.1, 20, No.10, pp.940-945 (1984)

[Oka 1987] 岡謙太郎："楽譜印刷", 共立出版, b i t 別冊, コンピュータと音楽 (1987)

[Oka 1990] 岡謙太郎："SGMLの音楽への応用", 音楽音響研究会資料, MA90-8 (1990)

[Ohteru 1985] 大照他："楽譜の自動認識システム (WABOT-2の視覚系)", 日本ロボット学会誌, Vol.3, pp. 153-162 (1985)

[Ohteru 1989] Ohteru, S. and Hashimoto, S. : "A New Approach to Music through Vision", Proc. Intl. Workshop on Artificial Intelligence and Music, pp.1-10 (1989)

[音楽の友社 1975] :新音楽辞典 (1975)

[Riecken 1989] Riecken, R. D. : "A K-line Model of Musical Creativity : Musical Composioton by Emotional Computaion", Proc. Intl. Workshop on AI and Music, pp.88-94 (1989)

[Roads 1985a] Roads, C. : "Research in Music and Artificial Intelligence", ACM computing Surveys, pp163-190 (1985)

[Roads 1985b] Roads, C. : Composers and the Computer, William A.Kaufmann (1985)

[Rodet 1988] Rodet, X. and Eckel,G. : "Dynamic Patches : Implementation and Control in the Sun-Mercury Workstatsion", Proc. ICMC, pp.82-89 (1988)

[Rosenthal 1988] Rosenthal, D. : "A Model of the Process of Listening to Simple Rhythms", Proc. ICMC, pp.189-198 (1988)

[Saiwaki 1989] Saiwaki, N., Katayose, H. and Inokuchi, S. : "An Approach to a Computer Assisted Composition System with KANSEI Parameter", Proc. 2nd. Intl. Workshop in AI and Music (1989)

[Seashore 1938] Seashore, C.E. : Psychology of Music, McGraw-Hill (1938)

[Senju 1987] Senju,M. and Ohgushi, K. : "How are the players ideas conveyed to audience?", Music Perception, 4, pp.311-323 (1987)

[Schenker 1956] H. Schenker : Der Freize Sats, Vienna Universal Editon (1956)

[Shimoda 1982] 下田和男：楽典, pp.85-104, 共同音楽出版社 (1982)

[Smolian 1980] Smolian, S.: "A Computer Aid for Schenkerian analysis", Compter Music Journal,

4, 2, pp.41-59 (1980)

[Takami 1989a] 高見啓史：楽曲分析能力をもつ自動演奏システム，大阪大学基礎工学部修士論文 (1989)

[Takami 1989b] 高見, 片寄, 井口："ピアノ演奏における演奏情報の抽出", 信学論, D-, Vol. J72-D-2, pp917-926 (1989)

[Taguti 1989a] 田口友康："自動演奏と演奏モデル", 共立出版, b i t 別冊, コンピュータと音楽 (1989)

[Taguti 1989b] Taguti, T. and Ohta, M. : "Agogics as a Metrical Rhythm and Pgrasing", Proc. ICMPC, pp.219-214 (1989)

[Tanguiane 1989] Tanguiane, A. S. : "A principle of Relativity of Perception and its Applications to Pattern Recognition in Analysis of Performed Music, Proc. ICMPC (1989)

[Tenny 1980] Tenny, J. and Polansky, L.: "Temporal Gestalt Perception in Music", Journal of Music Theory, 24(2), 205 (1980)

[Toda 1985] 戸田正直：心を持った機械, ダイヤモンド社 (1985)

[Tomotsuka 1988] 富塚, 白川, 五十嵐："音楽情報処理システムPSHYCHEと一,二の試み", 人工知能学会全国大会論文集, pp389-392 (1988)

[Uchio 1989] 内尾, 樋口, 北橋, 真田, 手塚："毛筆楷書文字の感覚的大きさの一正規化手法", 信学論, Vol. J72-D-2, pp1650-1656 (1989)

[Vercoe 1979] Vercoe, B. et al. : "MUSIC11 Manual", Experimental Music Studio, Cambridge, MA (1979)

[Vercoe 1985] Vercoe, B. et al. : "Synthetic Rehearsal" Training the Synthetic Performer" Proc. ICMC, pp.275 (1985)

[Vercoe 1988] Vercoe, B. and Cumming, D. : "Connection Machine Tracking of Polyphmnic Audio", Proc. ICMC, pp.211-218 (1988)

[Watkins 1985] Watkins, A. J.: "Perceptual Aspects of Synthesized Approximations to Melody", J.Acoust. Soc. Am. 78 (4), pp. 1977-1186 (1985)

[Winograd 1968] Winograd, T. : "Linguitics and the Computer Analysis of Tonal Harmony", J. Music Theory (1968)

(注)

ICMC = International Computer Music Conference

ICMPC = International Confernece on Music Perception and Cognition

付録1 評価に使う形容詞対

1. やばったい感じ	<-->	しゃれた感じ
2. ありきたりな感じ	<-->	特色のある感じ
3. 風変わりな感じ	<-->	型にはまった感じ
4. のどかな感じ	<-->	緊迫した感じ
5. 激しい感じ	<-->	穏やかな感じ
6. 堅い感じ	<-->	柔らかな感じ
7. はりつめた感じ	<-->	ゆったりした感じ
8. 男性的な感じ	<-->	女性的な感じ
9. 力強い感じ	<-->	弱々しい感じ
10. 若々しい感じ	<-->	老けた感じ
11. テンポの遅い感じ	<-->	テンポの速い感じ
12. のんびりした感じ	<-->	せわしい感じ
13. 変化にとんだ感じ	<-->	単調な感じ
14. 生き生きした感じ	<-->	生気のない感じ
15. 不安定な感じ	<-->	安定した感じ
16. 賑やかな感じ	<-->	落ち着いた感じ
17. 動的な感じ	<-->	静的な感じ
18. クールな感じ	<-->	ホットな感じ
19. うきうきした感じ	<-->	しみじみとした感じ
20. 沈んだ感じ	<-->	陽気な感じ
21. 軽やかな感じ	<-->	重々しい感じ
22. 古風な感じ	<-->	モダンな感じ
23. 進歩的な感じ	<-->	保守的な感じ
24. 明るい感じ	<-->	暗い感じ
25. 憂いをおびた感じ	<-->	晴れやかな感じ
26. 哀調をおびた感じ	<-->	明朗な感じ
27. ユーモラスな感じ	<-->	きまじめな感じ
28. おどけた感じ	<-->	深刻な感じ
29. 開放された感じ	<-->	抑圧された感じ
30. さっぱりした感じ	<-->	ねっとりした感じ
31. 窮屈な感じ	<-->	のびのびした感じ
32. ウエットな感じ	<-->	ドライな感じ
33. 派手な感じ	<-->	地味な感じ
34. 甘い感じ	<-->	しぶい感じ
35. がさつな感じ	<-->	優雅な感じ
36. うすっぺらな感じ	<-->	深みのある感じ
37. 俗っぽい感じ	<-->	高尚な感じ
38. 知性的な感じ	<-->	知性を欠いた感じ
39. 貴族的な感じ	<-->	庶民的な感じ
40. 素朴な感じ	<-->	洗練された感じ
41. 冷たい感じ	<-->	温かい感じ
42. 健全な感じ	<-->	退廃的な感じ
43. 複雑な感じ	<-->	単純な感じ

付録2 因子分析法

因子分析法は量的に測定されている項目を代表する総合指標を求める方法である。以下、因子分析手法について示す。

表1

個体 \ 変量	z_1	z_2	...	z_p
1	z_{11}	z_{21}	...	z_{p1}
2	z_{12}	z_{22}	...	z_{p2}
:	:	:		:
n	z_{1n}	z_{2n}	...	z_{pn}

表1のような p 変量 - n 個体の観測が得られたとする（本論文では、個体とは形容詞対、変量とは音楽プリミティブ、 z が評定値）。これら p 個の変量間の相関ができるだけ少ない共通因子を求めるのが因子分析法である。いま m 個の因子を考えて次のようなモデルを想定する。

$$\begin{aligned} z_{1i} &= a_{11} f_{1i} + \cdots + a_{1m} f_{mi} + e_{1i} \\ z_{2i} &= a_{21} f_{1i} + \cdots + a_{2m} f_{mi} + e_{2i} \\ &\dots \\ z_{pi} &= a_{p1} f_{1i} + \cdots + a_{pm} f_{mi} + e_{pi} \\ (i &= 1, 2, 3 \dots n) \end{aligned} \quad (1)$$

これを行列表現すると

$$z_i = A f_i + e_i \quad (i = 1, 2, 3 \dots n) \quad (2)$$

ここで A は因子負荷行列、 f_i は因子得点ベクトル、 e_i は独自因子の得点ベクトルと呼ばれる量である。共通因子はそれぞれ平均 0、分散 1、独自因子 e_1, \dots, e_p は平均 0、分散 d_1^2, \dots, d_p^2 、独自因子相互間および独自因子と共通因子間は互いに無関係と仮定する。

（共通因子 f_1, \dots, f_p 間の相関について互いに無相関と仮定する場合を直交因子と呼ぶ。）

直交因子の場合、式(2)のモデルが成り立つときには、 p 変量ベクトル Z の相関行列は

$$\Sigma = AA' + D \quad (3)$$

のように表される。ここで $D = d_i \cdot a g (d_1^2, \dots, d_p^2)$ である。よって表5-1のような Z の観測値に基づいて式(3)を満たす因子負荷量 A や独自因子の分散 D を求めることができる。

式(3)の対角要素は

$$l = \sigma_{jj} = h_j^2 + d_j^2 \quad (4)$$

$$h_j^2 = a_{j1}^2 + a_{jm}^2$$

と表され、 h_j^2 は共通因子 f による変動を表し、共通性と呼ばれている。

式(3)の関係より、相関行列の対角要素を共通性で置き換えた行列

$$\Sigma^* = \begin{bmatrix} h_1^2 & \rho_{12} & \dots & \dots & \rho_{12} \\ \rho_{21} & h_2^2 & \dots & \dots & \rho_{2p} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \rho_{p1} & \rho_{p2} & \dots & \dots & h_p^2 \end{bmatrix}$$

は共通因子によって説明される部分の分散、共分散を表し、もしモデルがよくあてはまっているならば、この Σ^* は階数の m の行列 A と A' に分解される。このとき Σ^* の階数も m になり、なりその p 個の固有値は m 個が正、 $p-m$ 個が0となる。固有値を $\lambda_1, \dots, \lambda_m$ 、これに対応する固有ベクトルを c_1, \dots, c_m ($\|c_j\|=1$ と正規化) とすると

$$\Sigma^* = \lambda_1 c_1 c_1' + \lambda_2 c_2 c_2' + \dots + \lambda_m c_m c_m'$$

よって(3)より

$$A = [\sqrt{\lambda_1}c_1 \ \sqrt{\lambda_2}c_2 \ \dots \ \sqrt{\lambda_m}c_m]$$

が求まる。

このようにして因子負荷行列Aが算出され、その要素の符号と大きさによって因子の解釈を行うことができる。

付録3 音楽プリミティブ（コード進行）

ここでは、音楽プリミティブのうちコード進行を取り上げて因子軸を求めるにした。概念空間を定義するという目的に限れば、すべての音楽プリミティブに対しての実験を行なう必要はない。下記のコード進行は、ポピュラー音楽で使われるコード進行のパターンのうち、頻度の高いものを8つのシーケンスとして並べたものである。

1. / Am / D / G / C / F / Dm / E7 / E /
2. / G / C / F / Bb / Eb / A / Dm / C /
3. / G / C / B m / Em / Am / D7 / G / G /
4. / C / G / Am / Em / F / C / F / C /
5. / C / G / C / F / C / Am / Dm / G /
6. / C / C / B / B / Bb / E / A / D /
7. / Am / G / F / Am / C / D / E7 / E /
8. / Am / G / F / Em / Dm / Em / Am / Am /
9. / Am / Dm / F / Em / Am / Dm / F / Em /
10. / F / C / Dm / C / F / C / Dm / C

付録4 評定値

形容詞対

コード進行

No.	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9	No.10
1.	8	7	5	6	4	8	5	4	6	4
2.	6	7	5	4	4	8	6	4	4	4
3.	7	4	6	6	6	3	6	6	6	6
4.	4	3	4	4	3	5	7	4	6	4
5.	6	8	6	6	7	6	4	6	6	6
6.	6	7	6	6	6	6	4	5	4	6
7.	6	7	6	6	7	6	4	5	4	6
8.	6	4	5	4	4	7	4	5	5	4
9.	6	5	5	5	4	7	4	5	4	5
10.	4	4	5	5	4	6	4	6	6	5
11.	5	3	5	5	4	6	6	4	4	5
12.	4	3	4	5	4	4	6	4	5	5
13.	4	6	6	5	6	6	4	6	6	6
14.	4	4	5	4	4	6	4	6	6	4
15.	4	4	6	6	6	3	4	6	4	7
16.	4	3	4	4	3	4	5	5	5	4
17.	6	7	5	5	7	6	5	7	6	6
18.	6	4	6	5	7	7	3	7	6	5
19.	4	4	4	5	5	3	6	4	3	5
20.	6	6	5	5	5	6	5	7	6	5
21.	4	6	5	6	5	4	6	3	3	5
22.	4	3	5	5	5	4	5	7	6	5
23.	7	7	5	5	3	6	5	3	6	4
24.	6	4	5	5	7	4	5	7	5	6
25.	4	6	6	6	6	3	4	3	4	6
26.	4	6	6	6	6	3	4	3	4	6
27.	6	5	6	5	6	5	6	7	6	6
28.	6	5	5	5	5	6	6	6	6	5
29.	5	3	4	5	5	4	6	6	6	5
30.	6	3	4	4	5	4	5	6	6	4
31.	5	8	6	5	6	5	4	5	5	5
32.	4	6	6	5	6	4	5	3	3	5
33.	5	6	6	5	6	7	4	7	7	7
34.	4	4	4	5	6	8	6	7	7	7
35.	6	6	6	5	6	7	4	6	6	5
36.	6	4	6	5	5	7	5	7	6	4
37.	6	5	6	5	6	6	4	6	6	4
38.	4	5	5	5	5	4	6	5	4	6
39.	4	5	5	5	6	5	5	4	5	6
40.	7	8	6	5	4	8	6	3	6	3
41.	4	6	7	6	6	4	4	4	4	6
42.	6	4	4	4	3	8	4	4	6	3
43.	4	6	5	5	6	6	6	5	8	

付録5 ショパン作曲：小犬のワルツ（中間部）

The musical score for 'Little Dog Waltz' (中间部) by Chopin, featuring six staves of piano music. The score includes dynamic markings such as *sostenuto*, *mf*, and *dolce*. Fingerings (1 through 6) are indicated above the notes, and 'T.A.' markings are placed below the notes. The key signature changes frequently, including sections in C major, G major, and E major.

付録6 メトロノーム速度

一般にメトロノーム数とは、単位音価を持つ音符が1分間打たれる個数（速度）を意味する。例えば、 $\text{♩} = 60$ という例では、1分間に 60 拍の早さで演奏すべきことを示している。メトロノーム数は普通楽節や楽章のように比較的長い区間に對して与えられる量である。このメトロノーム数を個々の音など局所的な区間に適用することを考える。いま、単位音価を s_0 とし、下層的に無限に小さい音価 ds を持った演奏時間を dt [min] とする。ここで、

$$m(s) = \frac{1}{s_0} \frac{ds}{dt}$$

とするような s の関数 $m(s)$ を局所メトロノーム速度と定義する。すなわち、実際の演奏における発音時刻 t_i が、

$$t_1, t_2, t_3, \dots, t_n, t_{n+1}$$

であったとするとき、音価 k を持った音符 i の単位音価 k_0 に於けるメトロノーム値は次のように定義される。

$$m_i = \frac{1}{k_0} \frac{k_i}{d_i}$$
$$d_i = t_{i+1} - t_i \quad (i = 1, 2, 3, \dots, n)$$