

音楽感性空間における非線形判別分析を用いた 曲印象別グループの分割

坂本 崇^{†,☆} 梶川 嘉延[†] 野村 康雄[†]

本論文では、非線形判別分析を用いて、計算機に与えた楽譜情報からその曲の特徴や印象を自動的に抽出し、任意のグループに分割する手法を提案する。特定の演奏者による演奏の個性や芸術性を抽出する際、情報处理的な観点からのアプローチを行うとして、演奏者がどのように楽譜を解釈し、演奏に至るかという命題の解決策は、(1) 楽譜に記述されている情報や音楽理論を考慮したプロダクションシステムによるルール生成法、(2) ニューラルネットワークを用いた解法、(3) 重回帰分析を用いた解法などが報告されている。これらの一問題点として、演奏者は楽曲の特徴や印象に応じて奏法を変える事実が存在するにもかかわらず、処理を行ううえで明示的に考慮されたものが存在しないことがあげられる。本論文ではその問題点を解決すべく、まず、因子分析やSD法を用い、実際のリスナーの立場に立った心理聴取実験を行う。そして、ニューラルネットワークによる非線形判別分析を用い、その楽曲はどのような印象を持つかをコンピュータが自動的に判別し、その印象にふさわしいグループに振り分ける手法を提案する。なお、このシステムでは未知の楽曲においても88%の精度でその曲印象を判別させることに成功した。

Classification of Group According to Tune's Impression Using Nonlinear Distinction Analysis in the Music Sensibility Space

TAKASHI SAKAMOTO,^{†,☆} YOSHINOBU KAJIKAWA[†]
and YASUO NOMURA[†]

In this paper, we propose a method which can extract tune's characteristic and impression automatically and classify into the arbitrary groups only from score information given to computer using the nonlinear distinction analysis. In these researches, the following methods are reported as a solution of the proposition how the player interprets a score, and gives the performance; (1) The rule generation method by production system considering the information described in the score and musical theory. (2) The method using neural network. (3) The method using multiple regression analysis. However, in spite of the fact that the skilled player changes his or her playing style according to the characteristic and impression of the tune, the system considering the fact has not existed yet. Therefore in this paper, in order to settle up the problem, the psychological listening experiment based on the standpoint of actual listeners is done using the SD method and the factor analysis at first. And applying the nonlinear distinction analysis by the neural network to the result, the computer automatically distinguishes what impression the tune has, and distributes the tune to a suitable group according to the impression. Furthermore, in an unknown tune, this system succeeded in the distinction of the impression of the tune by high accuracy of 88%.

1. はじめに

現在、さまざまな音楽制作の局面においてコンピュータ、シーケンサによる“打ち込み”が用いられていることは周知の事実である。1980年初頭にMIDI (Musical Instrument Digital Interface) と呼ばれるシンセサイ

ザー、サンプラーなどのデジタル機材を接続して使用するための国際的な統一規格の発足を皮切りに、音楽とコンピュータの関係は飛躍的に密接な関係となった。また、これらは一部のプロフェッショナルたちの独占物ではなく、だれでもが気軽にコンピュータを用いた音楽制作、鑑賞をすることを可能とした。しかしながら、まだ真に創造的な形での応用の環境は十分に整ってはおらず、あくまでも今日の音楽分野におけるコンピュータの利用は1つの手段にしかすぎない。現在広く用いられている音楽制作ツールは、ユーザの入

[†] 関西大学工学部
Faculty of Engineering, Kansai University
[☆] 現在、ヤマハ株式会社
Presently with YAMAHA Corporation

力に対し忠実にそれを再現してくれる反面、芸術性に富んだ音楽に不可欠な要素である感情のこもった表現をコンピュータ自ら発することはけっしてありえない。

これらの背景からも、音楽情報処理の分野において、特定のピアノ演奏者による演奏技術や芸術性、弾き癖といった人間の演奏時における感性を計算機によって抽出する研究も多く行われている^{1)~5)}。特定の演奏者による演奏の個性や芸術性を抽出する際、情報処理的な観点からのアプローチを行うとして、演奏者がどのように楽譜を解釈し、演奏に至るかという命題の解決策は、(1) 楽譜に記述されている情報や音楽理論を考慮したプロダクションシステムによるルール生成法^{2),6)}、(2) ニューラルネットワークを用いた解法³⁾、(3) 重回帰分析を用いた解法^{4),5)}などが報告されている。(1)においては、楽譜情報と、その楽譜情報における演奏を if~then 形式のプロダクションルールによって結び付ける手法である。しかし、この手法においては、演奏者が意識的に付加する解釈や弾き癖のルール化やその解釈、また、システムにとって既知である曲を再現することには有効であるが、演奏者の無意識のうちに付加される解釈や弾き癖のルール化や、システムの未知の曲に対する柔軟性に問題がある。(2)においては、非線形事象に対応し、未知曲の再生にも効果を発揮するが、処理の内部がユニット間の結合係数で表現されるため、ブラックボックス化されており、データ解析が困難である。(3)は楽譜情報を説明変数として表現し、重回帰分析における結合係数から演奏形態を決定するものであるが、これも基本的には(1)に帰着し、ルールベースにあたる説明変数の模索の必要性があり、未知曲への柔軟性に対する問題点があげられる。いずれの方法にも一長一短があり、トータル面で評価するとどの方法が秀でておりどの方法が劣っているという議論はナンセンスで、むしろ使用目的によるファクターが重要である。

しかしながら、(1)~(3)のいずれの方法にしても、処理を行わせる楽曲の選択方法は現在人間が任意に選び、また、選ばれた楽曲もすべて同一のフィールドで処理を行わせているものばかりである。しかしながら、演奏者は楽曲の特徴や印象に応じて奏法を変える事実が存在するにもかかわらず、処理を行ううえで明示的に考慮されたものが存在しない問題点があげられる。これは、処理を行うシステムがその曲の持つ印象を理解するモジュールを有していないためである。このモジュールをシステムに装することにより、明示的にその曲の持つ特徴や印象にふさわしいグループごとに処理を行うことができ、未知の楽曲においても、その曲

の特徴や印象にふさわしい演奏を行わせることが可能となる。また、グループ分類には Dannenberg ら⁷⁾の研究がある。しかし、この研究では演奏スタイルのパターン認識を行っており、楽譜情報から曲印象によるグループ分類を行うことは不可能である。そのため、その方法を自動演奏システムに装することは困難である。

そこで本論文では、自動演奏システムに装するための曲印象自動判別モジュールの手法として、重回帰分析や重判別分析を用いた線形的予測法から提案を行い、その限界を確認したうえでニューラルネットワークを用いた非線形判別分析に処理を拡張し、学習曲にとどまらず未知曲においても高精度に曲印象を自動判別した実験結果を示す。

2. 曲印象を考慮した自動演奏システムの構成と曲印象自動判別モジュールの生成方法

2.1 演奏者情報

本研究において演奏者情報とは、演奏者が実際に演奏を行った際に得られる情報で、かつ、その情報を扱うことによって演奏者の特徴を抽出することができるパラメータである。本研究ではこれらの情報を標準 MIDI ファイル (SMF) データを基にデータベース化している。現在用いている演奏者情報を表 1 に示す。

5 の Tempo とは、1 分間あたりの拍数 (単位は拍/分) である。ただし本論文では、実験に用いたすべての曲は 4 分音符を 1 拍の単位に変換し、処理を行っている。

2.2 印象別グループ分割の必要性

本手法の必要性を説明するにあたり、演奏者の特徴を if~then 形式によるエキスパートシステムを用いて抽出する場合を例にとる。表 2 は、処理単位 (ここ

表 1 演奏者情報

Table 1 Player information.

1. Velocity	: 鍵盤を押さえる強弱
2. Step Time(location)	: 鍵盤を押さえる絶対的時間
3. Gate Time(duration)	: 鍵盤を押さえている時間長
4. Note Number	: 押さえた鍵盤の名前 (音高)
5. Tempo	: 曲の流れる速さ

表 2 属性と演奏データ例

Table 2 Example of attributes and performance data.

Unit number	1	2	3	4
Chord	C	C	C	F
Ave. of note height	3	3	3	4
Note density	3	3	3	2
Velocity	60	64	66	72
Tempo	150	151	154	158

という処理単位とは、フレーズのおよその切れ目を実験者が手作業で区切ったものである)におけるコードネーム、平均音高、音符密度および、実際の演奏におけるヴェロシティ、テンポを分類したものである。また、平均音高と音符密度は5段階評価で1が最小としているが、これは実験に用いた楽曲のデータを総当たりし、平均的に5等分したものをを用いている。

たとえば、ルール抽出に用いる属性として、“コード”、“そのコード内における平均音高”および“音符密度”なるパラメータを用いた場合、条件部(LHS)に属性の過去1回の変化を考慮し、実行部(RHS)に前回の処理単位からの演奏データの差分を用いる場合、演奏データのうちヴェロシティに関しては、次の3つのプロダクションルールが得られる。

Rule 1

if (Chord is C → C(now),
Ave. of note height is 3 → 3(now)
Note density is 3 → 3(now))
then +4

Rule 2

if (Chord is C → C(now),
Ave. of note height is 3 → 3(now)
Note density is 3 → 3(now))
then +2

Rule 3

if (Chord is C → F(now),
Ave. of note height is 3 → 4(now)
Note density is 3 → 2(now))
then +6

ここでルール1とルール2は条件部がまったく同じになるので、実行部を平均することによって1つのルールにする(条件部平均化処理)ことにより、実行部の操作値は+3となる。

以上が、エキスパートシステムを用いた演奏者情報を抽出する一例だが、この例を見ても分かる通り、システムに学習させようとする楽曲を同一のフィールドで処理を行うと、生成されたルールは一般的なものになってしまう。しかしながら、熟練した演奏者は、曲の雰囲気、印象に応じた奏法を持っており、このように学習曲すべてに対し、同一部分による処理を行うと、その多様性が失われてしまう。たとえば、静かでスローな曲にはそれにふさわしい処理、激しくアップ

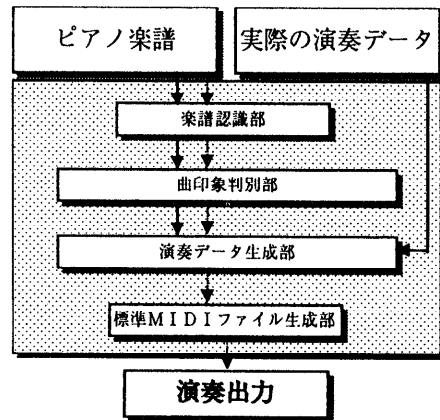


図1 曲印象を考慮した自動演奏システム
Fig.1 Automatic performance system considering tune's impression.

テンポな曲にはそれにふさわしい処理などである。これは、上に示したエキスパートシステムによる処理に限らず、ニューラルネットワークや重回帰分析などのアプローチにしてもまったく同じことである。したがって、学習曲の印象により自動的に分類し、その分類ごとに演奏者の特徴を抽出するモジュールを明示的に装している必要がある。

図1に曲印象による自動分割システムの構成図を示す。従来より自動演奏に関する研究は多くなされているが、演奏者は楽曲の特徴や印象に応じて奏法を変える事実が存在するにもかかわらず、処理を行ううえで明示的に考慮されたものが存在しない。そこで、本論文では曲印象を計算機が解釈し、その解釈された曲印象にふさわしいグループごとに演奏データを生成する手法を提案する。

2.3 心理聴取実験による音楽感性空間の生成

本節では、曲の表情別に学習曲を分類する^{6),8)}ために、まず予備的な実験としてSD法、ならびに因子分析⁹⁾による心理聴取実験を行い、その結果を三次元の因子空間として表現する。なお、この三次元の因子空間を本論文では音楽感性空間¹⁰⁾と呼ぶことにする。

実験条件は、聴取サンプルとして音楽演奏経験が十分にある、関西大学の交響楽団員40名と軽音楽部員40名の合計80名に対して行い、SD法の評価尺度は表3に示す7つの形容詞対を用い、それぞれ1~7の7段階で評価を行った。

また、被験者には「ブルグミュラー25の練習曲」から25曲を音楽大学講師による演奏で聴取させた。実験は、講師の演奏をYAMAHA P300で録音したカセットテープを会議室内でスピーカで呈示し、1曲が終了した時点で30秒以内に評価を行う。また被験者の中

表3 SD法で用いた形容詞対
Table 3 Pair of adjective (application in semantic differential method).

安定	⇔	不安定
滑らか	⇔	粗い
軽い	⇔	思い
柔らかい	⇔	硬い
透明	⇔	不透明
明るい	⇔	暗い
力強い	⇔	弱々しい

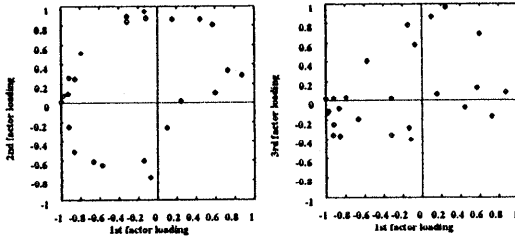


図2 因子分析結果
Fig.2 Result of factor analysis.

には、25の練習曲を知っている者もいるはずであり、順序効果を相殺するために、曲のタイトルを50音順にソートした順に呈示した。なお、交響楽団員40名と軽音楽部員40名の聴取実験は別の日に行っており、それぞれのデータを集計し、互いの相関係数を計算したところ、0.95を超えており、十分信頼の持てるデータである。

そして、これらの結果を因子分析し、解を第三因子までを求め、さらにローバリマックス回転を施した。このようにして得られた因子分析の解を図2に示す。

2.4 クラスタ分析による印象別グループの分割

図2は因子分析の解であるゆえ、曲どうしの相関係数がベクトルの内積で表現されている。その結果、点と点の距離が近ければ近いほど互いに同じ表情を持つ曲となる。逆にいえば、点と点が離れていれば離れているほど、互いの曲は反対の印象を持つ曲となる。そこで、これらの空間を任意のグループに分割するために、統計的手法であるクラスタ分析^{9),11)}を施した。図3にその結果の樹状図(dendrogram)を示す。

ここで、分割するグループ数について考察すると、図3における横軸(距離)はおよそ1.8から3.6の間が最も離れており、この間を切るような線を引き、それに従ってグループを分割するのがよいので、分割グループ数は3となる。図4は、感性空間を3つに分割し、それを三次元図上で眺めた結果である。

図4を見れば、視覚的にも理解しやすい形でグループが分割されたことが分かる。3章では図4の結果を

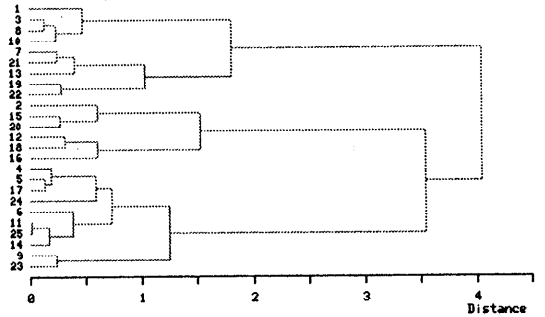


図3 クラスタ分析結果
Fig.3 Result of cluster analysis.

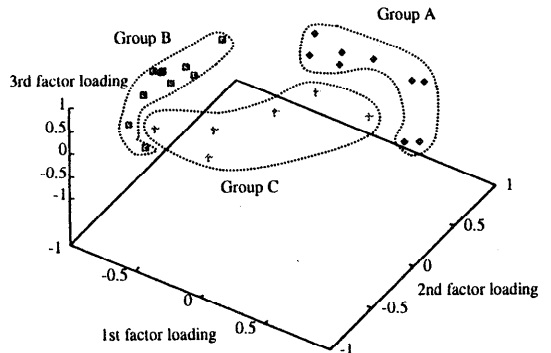


図4 グループ分割結果
Fig.4 Result of group division.

用いて、未知曲にも対応したグループ判別的手法を提案し、実験を行う。

3. 重回帰分析、重判別分析を用いた線形的なグループ分割法

2.4節では既知の楽曲に対して三次元空間を構成したが、自動演奏システムに装する以上、システムに未知である楽曲をもその印象別グループに特定する必要がある。そこで、本章ではまず、重回帰分析^{4),9)}、重判別分析を用いて、与えられた楽曲がどのグループに属するのかを線形的に予測する手法の提案を行い、さらにその実験結果を述べる。

一般に重回帰分析とは、 $n \times (p + 1)$ 個の値 $x_{i1}, x_{i2}, \dots, x_{ip}$ ($i = 1, 2, \dots, n$) が与えられたとき、近似式 $y_i \simeq f(x_{i1}, \dots, x_{ip})$ を満たす関数 f を求めるものである。 $p = 1$ を単回帰分析といい、 $p > 1$ を重回帰分析という。また、 y_i を目的変数といい、 $x_{i1}, x_{i2}, \dots, x_{ip}$ を説明変数という。すなわち、

$$y_i = b_1 x_{i1} + b_2 x_{i2} + \dots + b_p x_{ip} + e_i \quad (1)$$

とおき、 e_i (残差) の2乗和 $e_1^2 + e_2^2 + \dots + e_n^2$ が最小になるような b_1, b_2, \dots, b_p 、すなわち結合係数を求

める方法である。

本章では、重回帰分析と重判別分析を用い、線形的に印象別グループを特定する手法を提案する¹¹⁾。まず、式(1)を用い、説明変数に感性空間座標値(第1因子, 第2因子, 第3因子), 目的変数に楽譜情報(楽譜や演奏者情報から得られるペロシティやゲートタイム, ステップタイムの平均, 分散, 標準偏差や, テンポ, 調性など)を与え, 結合係数を求める。次に未知曲認識時においては, さきほど得られた結合係数を説明変数に設定し, 目的変数に楽譜情報を入力し, その結合係数を求める。その結果得られた結合係数が, 感性空間座標値となる。

以上のことを式を用いて説明する。まず, 最初の重回帰分析は $p = 3, n = 25$ として,

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{25} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \vdots & \vdots & \vdots \\ x_{251} & x_{252} & x_{253} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_{25} \end{bmatrix} \quad (2)$$

の残差二乗和 $e_1^2 + e_2^2 + \dots + e_{25}^2$ が最小となる結合係数 $[b_1 \ b_2 \ b_3]^T$ (T : 転置行列) を求める。このような結合係数をすべての楽譜情報について求める。つまり, 楽譜情報を3種類とした場合,

$$\begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}$$

が得られる。次に未知曲に対し, 3種類の楽譜情報を $[Y_1 \ Y_2 \ Y_3]^T$ とすると,

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} + \begin{bmatrix} E_1 \\ E_2 \\ E_3 \end{bmatrix} \quad (3)$$

とし, 残差二乗和 $E_1^2 + E_2^2 + E_3^2$ が最小となる結合係数を求めたとき, これから求める $[B_1 \ B_2 \ B_3]^T$ が感性空間座標値となる。

そして, 求められた座標値について重判別分析を施

表4 学習曲再生時におけるグループ予測正解率
Table 4 Probability of correctness (learning and recognition).

	曲数	正解数	正解率 [%]
グループA	9	6	66.7
グループB	10	7	70.0
グループC	6	4	66.7
合計	25	17	68.0

すと, その曲が属するグループが予測される。重判別分析とは, これも多変量解析の分野における一手法で, いくつかのグループに分類された対象についての関係を調べるためのものである。この方法を使えば, 新しい対象が与えられたとき, それがどのグループに属するのかを判別することができる。ここでいう重判別分析はマハラノビスの汎距離に基づいている。

表4は, 以上の手順により, 線形的に曲印象グループを予測した結果である。実験に用いた楽譜情報は, 1曲中に存在するメジャーコードの時間的な割合(調性), 楽譜に記述されているテンポを4分音符単位に変換した値(平均テンポ), 個々の音符が何分音符か, という情報を1曲中で眺め, それらの分散値(音符長分散)を求めた合計3種類のデータである。なおこの結果は, 判別させる曲に対し, その結合係数を求めるときにその曲の情報も含んでいるので, いわゆる“学習曲再生”になる。

このように学習曲再生で32%の誤認識をする要因として,

- 1) 予測変数が処理を行ううえで有効な情報を持っていない,
- 2) 予測変数間の非線形な関係,

の2点が考えられる。1)に関しては1曲の楽譜情報から得られる, 音符長, コード, テンポ, 音高の統計学量(平均, 分散, 標準偏差など)について求め, それらを重回帰分析にかけて予測したが, 表4以上の正解率をあげるものはなかった。よって, 本章で行った線形的な予測法の限界を確認したうえで, 4章では2)の予測変数間の非線形な関係に注目し, ニューラルネットワークを用いた非線形処理による印象別グループを予測する手法を提案する^{12), 13)}。

4. ニューラルネットワークを用いた非線形的なグループ判別法

4.1 線形から非線形への拡張

本研究で用いたニューラルネットワークは, フィードフォワード結合を持つ階層型のもので, 学習アルゴリズムは誤差逆伝搬法を用いた。入力データ, 教師デー

タはすべて 0.05~0.95 の範囲で正規化されたものを用いた。また、中間層は 2 層あるいは 3 層のものを用いた。

ところで、式 (1) はいわゆる線形結合の式であるが、これは一種のニューラルネットワークとも解釈できる。すなわち、式 (1) は、中間層のないニューラルネットワークと等価となり、ニューラルネットワークを用いることは、線形的に予測を行ってきた場合と本質的には変わらなく、線形処理をさらに非線形処理に拡張したと考えるべきである。

4.2 印象別グループ予測実験 [1]

まず、3 章で用いた目的変数 (楽譜テンポ、調性、音符分散) の 3 種類をニューラルネットワークの入力信号に用い、グループ予測を行った。なお、先にも述べたが、この 3 つの入力信号は、線形予測を行った際に最も予測結果が良かったものである。ニューラルネットワークの学習条件としては、

- 入力信号：楽譜テンポ、調性、音符長分散の 3 ユニット、
- 教師信号：感性空間座標値、
- 中間層：10-10、

とし、エラーポテンシャルが $1e-6$ 未満になった以降に学習を終了する。なお、本論文中で中間層の規模を 10-10 と記述した場合は、中間層が 2 層で、その各層が 10 ユニットからなることを意味する。以上の条件で、感性空間座標値を予測し、それらを重判別分析にかけ、グループを予測した結果を表 5 に示す。なお、ニューラルネットワークの学習が終わった時点で、再び入力ユニットに学習に用いた楽譜情報を入力し、ネットワークに認識を行わせた結果 (学習曲再生)、100% の正解率となり、線形予測の限界を非線形処理で補うことができた。さらに、表 5 は、ターゲットとなる楽曲を除いた 24 曲を学習させ、その残り 1 曲を未知曲として扱い、ニューラルネットワークに認識させた結果を 25 パターン集め、正解率を求めたものである。

学習曲再生は、100% 正解するのは当然であり、未知曲についてもある程度認識できたことが分かる。しかし、3 章でも述べたように、楽譜情報の統計学量や

ニューラルネットワークの中間層の構成を変えて同様の実験を行ったが、表 5 を超える正解率を得ることができず、違う判別方法についての検討を行う必要がある。そこで、4.3 節では、別のアプローチでグループ予測実験を行い、その結果についても述べる。

4.3 印象別グループ予測実験 [2]

4.2 節では、これまで線形的にグループ予測を行ってきた場合とまったく同じ条件でニューラルネットワークを用いて実験を行い、その有用性を確認したが、ここではその問題点を考察する。それは、重判別分析において、マハラノビス変換を行った後の座標において、その分布が理想的な円形になっていないという点である。マハラノビス変換を行うと、理想的には、そのグループごとの分布が円形になり、新たに与えられた対象のグループを判別する際、各々の円の中心とその対象の点との距離を求め、最も近いグループが該当するグループとなる原理である。一般的に、重判別分析を行う基のデータが線形的に分離されるときにはマハラノビス変換もうまく円形に分布されるのだが、非線形的に分類されるときには十分に円形に分布されず、重判別分析の効果が発揮されない。ところが、我々の扱う感性空間は、特に線形的に分布される、といった条件はまったく入っていない。すなわち、重判別分析を用いることは有用でないと考えられる。実際に、マハラノビス変換を行った後の感性空間座標を図 5 に示す。この図を見ても円形に分布していないことが分かる。

そこで、本節では感性空間座標値を予測してから印象別グループを予測するのではなく、楽譜情報から直接グループを予測する実験を行い、その結果を示す。ニューラルネットワークの学習条件は、

- 入力信号：調性、音高偏差、楽譜テンポ、音符長平均、音高 5 ユニット、
- 教師信号：グループごとにユニットを設け (3 ユ

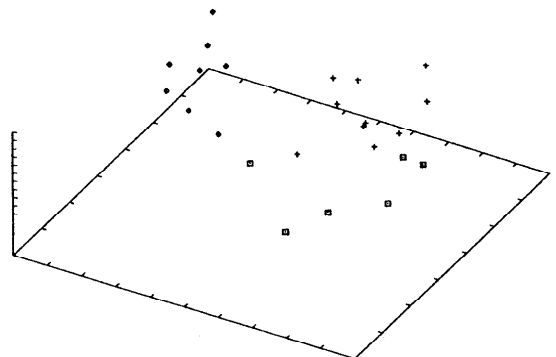


図 5 マハラノビス変換後座標

Fig. 5 The coordinate after Mahalanobis transformation.

表 5 未知曲認識時におけるグループ予測正解率
Table 5 Probability of correctness (unknown tune recognition).

	曲数	正解数	正解率 (%)
グループ A	9	7	77.8
グループ B	10	7	70.0
グループ C	6	4	66.7
合計	25	18	72.0

表6 未知曲認識時におけるグループ予測正解率
Table 6 Probability of correctness (unknown tune recognition).

	A	B	C	計	正解率 [%]
20-20-20	6/9	8/10	5/6	19/25	76
10-10-10	6/9	8/10	5/6	19/25	76
5-5-5	6/9	8/10	5/6	19/25	76
15-15-15	6/9	8/10	5/6	19/25	76
7-7-7	7/9	8/10	6/6	21/25	84
20-20	7/9	9/10	6/6	22/25	88

ニット), 該当するユニットに 0.95, その他のユニットに 0.05 を与える,

○ 中間層: 10-10-10,

とした。その結果を表6に示す。

このように、中間層 20-20の場合、未知曲における認識結果は 88%となり、非常に良い結果となった。そのうえ、少々中間層の数を変化させてもコンスタントに 76%以上の正解率を得ることができた。また、ニューラルネットワークの学習回数について考察すると、中間層 20-20-20, 10-10-10, 5-5-5, 15-15-15 の場合は約 6.6~8.2 万回とある程度近い値を出し、それぞれの出力結果もまったく同じのものであった。ところが、中間層 7-7-7では、学習回数が約 17400 回で、正解率も上がり、さらに、20-20では学習回数が約 14000 回と、これまでに比べ早く学習を行い、正解率も向上しており、中間層の設定はこの規模で行うのが良いことが分かる。

5. 感性空間座標と演奏パラメータとの関係

曲印象自動モジュールを生成するにあたり、4.3 節の実験結果では未知曲において 88%の精度でその曲の印象を判別することに成功した。本章では、感性空間座標と実際の演奏者の演奏データはどのような関連性があるのかを述べる。

ここで、図6に曲印象に応じた演奏データ生成の流れを示す。

本研究で望ましいのは、楽譜情報から直接感性空間座標値を求め、その座標値を参考にした演奏データ生成という流れである。しかしながら、4章では感性空間座標値を直接求めるのではなく、曲印象別のグループを予測している。そこで、この章では、楽譜情報と感性空間座標の予測ではなく、演奏情報と感性空間座標値の関連性について考察し、実験結果から本研究の有効性について確認する。

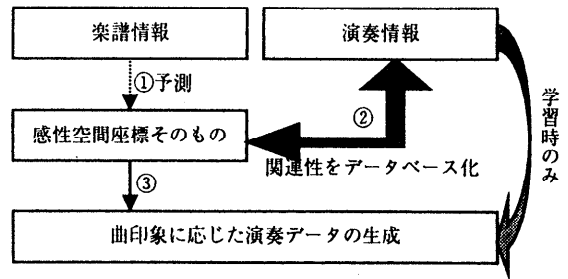


図6 曲印象に応じた演奏データ生成の流れ
Fig. 6 The flow of generating performance data according to each tune's impression.

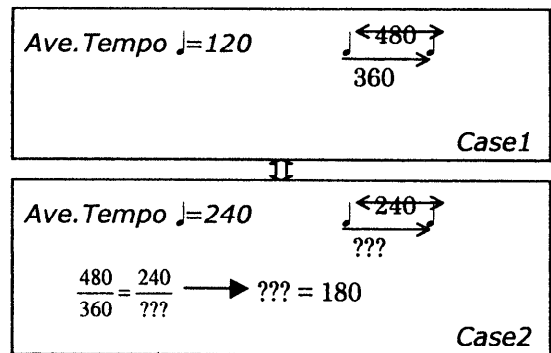


図7 曲のキレの良さ
Fig. 7 The significance of product of averaged gate time and averaged tempo.

実験アルゴリズムは、ニューラルネットワークを用いた非線形予測で、入力信号に感性空間座標値、教師信号に演奏データの各種パラメータを入力する。実験に用いた教師信号のパラメータは、演奏データから得られる、ゲートタイムの平均、分散、標準偏差、演奏テンポの平均、分散、標準偏差、ヴェロシティの平均、分散、標準偏差、そしてゲートタイム平均値とテンポ平均値の積の計 10 種類である。

ここでゲートタイム平均値とテンポ平均値の積は、「曲のキレの良さ」を表すパラメータである。その理由を図7で説明する。

たとえば、case1の手前の音符は理想的な値の 2/3 のキレを持っている。これは、平均で求められたテンポが 120 であるとき、平均的なゲートタイムは 360 である。また、テンポの単位は四分音符であるので、その四分音符の理想ゲートタイムが 480 である。そして、case2で考えると、平均テンポが 240 であるならば、同じキレを持つためには、2/3 である 180 がゲートタイムになる必要がある。また、このとき、もっとキレの良い曲ならば平均ゲートタイムは下がり、キレが悪かったら平均ゲートタイムは上がる。この平均ゲートタイムを、テンポの値に入れて考慮することにより、

表7 感性空間座標と演奏パラメータとの関係

Table 7 The recognition result of relation between sensibility space and performance information.

	相関係数		評価値	
	個別学習	一括学習	個別学習	一括学習
ゲートタイム 平均	0.281104	-0.100282	14.015053	14.41352
ゲートタイム 分散	0.403773	0.632387	17.341783	9.117913
ゲートタイム 標準偏差	0.479287	0.642024	22.501054	3.155043
テンポ 平均	-0.264839	-0.146653	21.340784	22.19559
テンポ 分散	-0.226954	-0.110420	227.08056	72.41508
テンポ 標準偏差	0.105772	-0.041770	3.376449	4.366917
ヴェロシティ 平均	0.809771	0.824599	1.998002	1.499584
ヴェロシティ 分散	0.582597	0.350793	27.83719	26.41764
ヴェロシティ 標準偏差	0.346229	0.318685	40.58056	30.25709
キレの 良さ	-0.215815	-0.271539	147.4706	93.84269

同一の条件で、曲のキレを判断することができる。

ここで、ネットワークの学習方法を2種類用意する。1つは「個別学習」で、教師信号のパラメータを1つずつ学習させるパターンで、当然出力ユニット数は1である。もう1つは「一括学習」で、教師信号のパラメータ10個をまとめて学習させる方法で、出力ユニット数は10である。

このような手順を用いて未知曲の認識を行った結果を表7に示す。なお、表中に与えた相関係数 r は式(4)、評価値 v は式(5)で求めた。

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \quad (4)$$

$$v = \sum_i \frac{(x_i - y_i)^2}{x_i^2} \quad (5)$$

この実験の場合、データ配列の大小変化の似かより具合を見る相関係数 r を用いるより、誤差を教師信号値で割る評価値 v を考察する方が妥当であるので、 r の値は参考程度である。評価値 v を見ると、まず全体的な傾向として一括学習を行った場合の方が個別学習を行った場合よりも誤差が小さい。そして、ゲートタイムの標準偏差における v は3.155043、ヴェロシティの平均値における v は1.499584であり、1曲あたりの平均の誤差は前者が0.1577、後者が0.07498程度であり、おおよそ感性空間と演奏情報の結び付きを再現できている。すなわち、本論文で提案する曲印象自動判別モジュールは、曲の揺らぎ具合を表すゲートタイムの標準偏差と、音量に対する演奏者のテンション(気分の高揚度)を表すヴェロシティの平均を抽出する

のに有効であると考えられる。しかしながら、今回導入した「キレの良さ」では良い結果が得られなかった。それは、今回用いたのは右手のキレの良さだけであり、左手のキレの良さが考慮されていなかったためと考えられる。すなわち、左手がよくキレていて、右手がレガート気味の演奏でも、今回の計算方法ならばキレの良さはレガート気味な値となるためである。また、その他のパラメータについては、ニューラルネットワークの規模を変化させても良い結果が得られなかった。つまり、感性空間からはそのような演奏情報との関連性を見いだすことができなかった。

6. まとめ

本論文では、まず重回帰分析、重判別分析などを用いた線形的な予測から、印象別グループの予測を行い、その限界を確認したうえで、ニューラルネットワークを用いた非線形判別分析を行うことによって、未知曲にも高精度なグループ予測を行うことに成功した。また、印象判別に用いた音楽感性空間と演奏者情報との関連性についても未知曲認識で十分な成果を得ることができた。音楽感性空間は楽譜情報と直接的なつながりを持たないが、演奏者は楽譜情報を独自に解釈し、それぞれの曲印象に応じた奏法を駆使することが示唆される。すなわち、「演奏者は曲の雰囲気に応じて奏法を変えて演奏を行う」といった仮説は実験結果からも有用であることが示唆される。したがって、このモジュールは作曲や編曲システム、そして、楽曲検索システムなどへの応用も考えられる。

今後の研究方針としては、音楽感性空間の空間座標値をストレートに、かつ高精度に予測する手法の模索、実験を行い、感性空間座標値からのグローバルな演奏者情報の抽出の研究も行う予定である。また、本論文では、入出力の因果関係がブラックボックスであるニューラルネットワークを使用しているが、本研究は計算機が曲印象を自動的に判別することを目的としており、ニューラルネットワークで模擬された因果関係を見つけることではない。しかしながら、ニューラルネットワークが模擬した入出力関係が明確化されれば、楽譜情報と曲印象との関係が分かり、非常に有益だと考えられる。そのため、ニューラルネットワークの入出力関係を明らかにすることも今後の研究課題としたい。さらに、本論文において説明変数(楽譜情報)の模索にはカットアンドトライ的な要素が非常に多く、定量的な結論を導き出すのは困難であった。しかし、回帰分析やニューラルネットワークで用いる入力変数のカットアンドトライについては、それを専門的に

う研究(データマイニングなど)を導入することにより, 定量的に説明変数の模索を行いたいと考えている。

謝辞 本研究の遂行にあたり, ご協力いただいた関西大学工学部電子工学科池添剛氏に感謝いたします。

参考文献

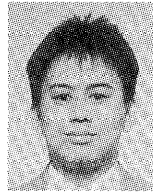
- 1) Widmer, G.: Understanding and Learning Musical Expression, *Proc. ICMC*, pp.268-275 (1993).
- 2) 白川健一, 小田安彦, 熊谷俊行, 梶川嘉延, 野村康雄: 演奏者情報を加味したピアノの自動演奏システムの構築—音符属性に着目した個人的解釈ルールの生成, 情報処理学会研究報告, MUS-10-1, pp.1-6 (1995).
- 3) 小田安彦, 白川健一, 村上 豊, 梶川嘉延, 野村康雄: 演奏者情報を加味したピアノの自動演奏システムの構築—ニューラルネットワークによる局所部分に於ける演奏特徴の抽出, 情報処理学会研究報告, MUS-10-2, pp.7-12 (1995).
- 4) 青野祐司, 片寄晴弘, 井口征士: 重回帰分析を用いた演奏表現法の抽出, 情報処理学会論文誌, Vol.38, No.7, pp.1473-1481 (1997).
- 5) 山内敏彰, 野池賢二, 野瀬 隆, 乾 伸雄, 小谷善行, 西村恕彦: GAを用いたピアノ演奏の局所テンポ解析, 情報処理学会研究報告, MUS-14-1, pp.1-6 (1996).
- 6) 白川健一, 熊谷俊行, 坂本 崇, 梶川嘉延, 野村康雄: 演奏者情報を加味したピアノの自動演奏システムの構築—曲の特徴を考慮したルールの生成, 日本音響学会平成8年度春季研究発表会講演論文集, 2-7-4, pp.667-668 (1996).
- 7) Dannenberg, R.B., Thom, B. and Watson, D.: A Machine Learning Approach to Musical Style Recognition, *ICMC97* (1997).
- 8) 坂本 崇, 藤井敬三, 波多正臣, 梶川嘉延, 野村康雄: 演奏者情報を加味したピアノの自動演奏システムの構築—曲の印象, 特徴による感性空間に関する検討, 日本音響学会平成8年度秋季研究発表会講演論文集, 2-5-18, pp.643-644 (1996).
- 9) 岡村晴彦: パソコンによるデータ解析入門, 技術評論社 (1986).
- 10) 片寄晴弘, 才脇直樹, 井口征士: 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講演論文集, 9-8, pp.453-456 (1989).
- 11) 坂本 崇, 波多正臣, 藤井敬三, 梶川嘉延, 野村

康雄: 曲印象を考慮したピアノの自動演奏システムの構築, 第54回情報処理学会全国大会論文集, 7J-3, pp.271-272 (1997).

- 12) 豊田秀樹: 非線形多変量解析, 朝倉書店 (1996).
- 13) 坂本 崇, 梶川嘉延, 野村康雄: 音楽感性空間を用いた非線形処理による曲印象や演奏者情報の自動抽出に関する検討, 第55回情報処理学会全国大会論文集, 1J-2, pp.2-3 (1997).

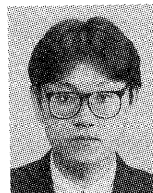
(平成10年2月12日受付)

(平成11年1月8日採録)



坂本 崇 (正会員)

平成8年関西大学工学部電子工学科卒業。平成10年同大学院博士課程前期課程修了。同年ヤマハ(株)入社。在学中, 音楽情報処理の研究に従事。



梶川 嘉延

平成3年関西大学工学部電子工学科卒業。平成5年同大学院博士課程前期課程修了。同年富士通(株)入社。平成6年関西大学助手。平成10年より同大学専任講師。主に, 電気音響変換器のCAD, 適応信号処理, 非線形信号処理の研究に従事。工学博士。電子情報通信学会, 日本音響学会, 電気学会, 計測自動制御学会, IEEE, EURASIP各会員。



野村 康雄 (正会員)

昭和36年大阪大学工学部通信工学科卒業。昭和38年同大学院博士課程修了。同年松下電器産業(株)入社。同社無線通信研究所勤務。昭和45年大阪大学研究生。昭和50年同大学助手。昭和51年関西大学講師。同助教授を経て, 現在, 工学部電子工学科教授。電気音響変換器のCAD, 人工知能, 音楽情報処理の研究に従事。工学博士。電子情報通信学会, 日本音響学会, IEEE各会員。