

実演奏音と楽譜情報からの演奏者識別に関する研究

中山 英樹[†] 大津 展之^{††}

[†] 東京大学大学院情報理工学系研究科, 〒113-0033 東京都文京区本郷 7-3-1

^{††} 産業技術総合研究所, 〒305-8568 茨城県つくば市梅園 1-1-1

E-mail: [†]nakayama@isi.imi.i.u-tokyo.ac.jp, ^{††}otsu.n@aist.go.jp

本研究の目的は、実演奏音から各演奏者の個性を抽出し、演奏者の識別を行うシステムの構築である。実演奏音の解析は技術上の困難が伴うため、演奏者の個性に関連した従来研究では、MIDI データをソースとしたものが多い。本研究では、市販のヴァイオリン独奏の CD をソースとし、実演奏音から演奏者の個性を評価する手法を提案した。まず、演奏者の個性を反映した特徴の抽出を行い、演奏特徴ベクトルを得た。これを線形判別分析によって分離することにより、確かに各奏者ごとに演奏が分かれることを示し、各奏者の定性的な傾向について議論した。また、その応用として未学習の曲の奏者を識別する実験を行い、良好な結果を得た。

Hideki NAKAYAMA[†] and Nobuyuki OTSU^{††}

[†] Graduate School of Information Science and Technology, University of Tokyo

7-3-1, Hongo, Bunkyo-Ku, Tokyo, 113-0033, Japan

^{††} National Institute of Advanced Industrial Science and Technology

1-1-1, Umezono, Tsukuba-shi, Ibaraki, 305-8568, Japan

E-mail: [†]nakayama@isi.imi.i.u-tokyo.ac.jp, ^{††}otsu.n@aist.go.jp

We developed a system to extract and learn music player's personality from real sound and to discriminate each player. Due to the difficulty of analyzing real sound, previous works adopt MIDI data as the source from which to extract music player's personality. We propose a method to evaluate music player's personality from real sound, and successfully applied it to the violin playings recorded in CDs on the market. First, the system extracted feature vector reflecting each players's personality and successfully separated each player in linear discriminant analysis. The system can also identify players from unknown playing.

1. はじめに

近年、音楽のコンテンツそのものに深く関わる音楽情報処理の研究が盛んになってきている。例えば、好みのタイプの楽曲を分類するシステム [1] や、サビ区間検出を行うシステム [2]、自動演奏に表情付けを行うシステム [3] など が挙げられる。

このようなシステムの実現のためには、人間の音楽理解の本質にアプローチする必要がある、そのような研究が新たな分野として注目を浴びている。その中でも、演奏者の個性(演奏スタイル)をモデル化する研究は、機械により人間らしい音楽を獲得させるために非常に重要と言える。演

奏者の個性に関わる研究ではこれまで、解析が容易である MIDI データをソースにすることが主流であった [4] [5]。しかし MIDI データは、あくまで通信データに過ぎず、電子楽器にしか適用できない。広範に適用できるシステムを作るためには、実音響をソースにすることが必須である。

本研究では、市販の音楽 CD 中の実音響データを直接解析し、各演奏者の個性を学習し、未学習データとして与えられた演奏を聴いて誰の演奏が識別するシステムを構築する。この際、楽譜情報を信号解析、特徴抽出の両場面において活用することを考える。題材とする曲は、J.S.Bach 作曲『無伴奏ヴァイオリンのためのソナタとパルティータ』とする。これは、バイオリン独奏のための曲集である。全

部で6つの曲からなり、多くの演奏家が録音を残している楽曲である。本研究では、バルティータ第1番第1楽章 (Double)、第2楽章 (Corrente)、第2楽章 (Double)、第3楽章 (Double)、第4楽章 (Double)、ソナタ第1番第4楽章 (Presto) を用いた。

2. アルゴリズム

本研究で提案するシステムの全体図を図1に示す。処理の流れは次の通りである。

- (1) 演奏データを、テンプレートマッチングにより、音一つ一つに分離
 - (2) 各音ごとの物理特徴を抽出
 - (3) 得られる物理特徴に楽譜情報を付加し、音情報ベクトルを得る
 - (4) 多数の音情報ベクトルの統計的傾向から、演奏情報ベクトルを得る
- 以下、順に説明する。

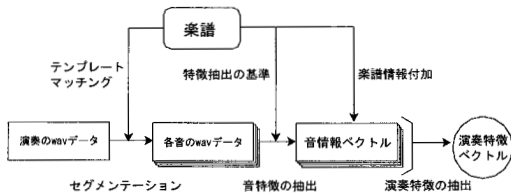


図1 システム全体図

2.1 音のセグメンテーション

単音レベルでの特徴を抽出するために、まず音のセグメンテーションを行う必要がある。ここでは、周波数領域でのテンプレートマッチングによりこれを行う。周波数解析には高速フーリエ変換を用い、標本点数は1024点(0.023sec)とした。

演奏時間と楽譜から、平均的な演奏テンポが計算できる。このテンポに従い、注目する音とその前後の計3音のテンプレートを作る。このテンプレートは、基本周波数の第5倍音までのパワー値を単純に加算するものである。

さらに、平均のテンポから、注目する音が時間軸上のどこに表れるのか大まかな位置が推定できる。この点の近傍をテンプレートマッチングで探索する(図2)。

この手法は演奏の途中でテンポが変動するとうまくいかないが、今回用いた曲では最初から最後までほぼ一定のテンポが保たれているため、30秒から1分程度の曲であればほぼ問題なく動作する。

2.2 各音の物理特徴量

各音から抽出する物理特徴量を表1に示す。以下、各特徴について順に説明する。

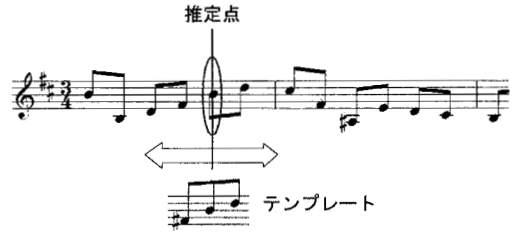


図2 テンプレートマッチング

表1 各音の物理特徴量

名称	説明
<i>Length</i>	音の長さの相対値
<i>Pitch</i>	音程の平均律からのずれ
<i>PitchWidth</i>	前音からの音程の間隔
<i>Power</i>	音の強さの相対値
<i>Stand</i>	音の立ち上がりの強さ

Length (音長)

ある音が立ち上がってから、次の音が立ち上がるまでをその音の継続時間 *duration* とする(図3)。これは、セグメンテーションされた一つ一つの音の長さそのものである。このままでは継続時間の絶対値に過ぎないので、「ずれ」の程度を表すようにするため、演奏全体を通じての平均値によって正規化を行う。演奏全体に含まれる音の数を N とすると、

$$Length = \log \frac{duration}{\frac{1}{N} \sum_{i=1}^N duration(i)} \quad (1)$$

Pitch (音程のずれ)

各音を、音の幅いっぱい (*duratoin*) を使ってフーリエ変換する。このようにフーリエ変換の標本点をできるだけ増やすことで、高精度の周波数分解能が期待できる。第 k 倍音の周波数を *frequency(k)* とする。これらからそれぞれ基本周波数を推定し、その平均を演奏音の周波数 *freq* とする。計算には、第5倍音まで用いた。

$$freq = \frac{1}{5} \sum_{k=1}^5 \frac{frequency(k)}{k} \quad (2)$$

さらに、これが平均律音程からどの程度離れているかをもって音程のゆらぎの特徴 *Pitch* とする(図3)。注目すべき音の平均律における周波数を *basefreq* とし、

$$Pitch = \log \frac{freq}{basefreq} \quad (3)$$

によって *Pitch* を定める。

PitchWidth (音程の間隔)

前音との音程の間隔を、ずれの差をとることで評価する。

$$PitchWidth = Pitch(t) - Pitch(t - 1) \quad (4)$$

例えば、注目する音が高く、前音が低くずれている時、音程の間隔は平均律よりも広がることになる。また、両方も同程度に高くずれている時は、音程の間隔は平均律と同程度となる。

つまり、この値が大きいくほど、前音との音程の間隔が平均律音程よりも広く、小さいほど狭いことになる。

Power (音の強さ)

各音を、音の幅いっぱいを使ってフーリエ変換し、第5倍音までのピークのパワー値を足し合わせたものを、演奏音の大きさ *loudness* とする (図 3)。これも *Length* 同様、平均によって正規化を行う。

$$Power = \log \frac{loudness}{\frac{1}{N} \sum_{i=1}^N loudness(i)} \quad (5)$$

Stand (音の立ち上がり)

Pitch や *Power* を求める際には、音の長さ全部を用いフーリエ変換を行ったが、ここでは音の出だしを細かくフーリエ変換することで、音の強さの立ち上がり (傾き) を求める (図 3)。ここでは、フーリエ変換の標本点を 2048 点 (0.046sec) とした。最初の 2048 点における音の *Power* 値を *pow1*、次の 2048 点における *Power* 値を *pow2* とした時、

$$Stand = \log \frac{pow2}{pow1} \quad (6)$$

によって定める。

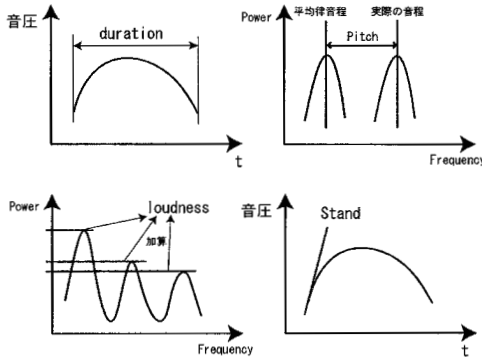


図 3 音から抽出する特徴

2.3 楽譜情報の付加

前節で求めた特徴は、単に各音の物理的特徴にすぎない。本節では、表 2 に示した楽譜情報を付加する。これらを図示すると図 4 のようになる。以下、それぞれ詳しく説明をする。

表 2 付加する楽譜情報

名称	説明
<i>Bar</i>	小節の頭の音であるか否か
<i>Beat</i>	拍頭の音であるか否か
<i>Height</i>	音符の高さ (音番号)
<i>Prev</i>	前の音からの音高の変化
<i>Next</i>	次の音への音高の変化

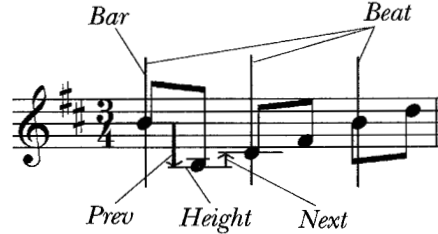


図 4 付加する楽譜情報

Bar (小節頭)

その音が、小節頭の音であれば 1、そうでなければ 0 の関数値である。

$$Bar = \begin{cases} 1 & \text{小節頭の音であるとき} \\ 0 & \text{それ以外のとき} \end{cases} \quad (7)$$

Beat (拍頭)

その音が、拍頭の音であれば 1、そうでなければ 0 の関数値である。

$$Beat = \begin{cases} 1 & \text{拍頭の音であるとき} \\ 0 & \text{それ以外のとき} \end{cases} \quad (8)$$

Height (音高)

音の高さを数値で表したものである。音が半音高くなるごとに 1 値が増える。今回は、平均律音程で 130.8Hz の C(ド) を 0 としたので、この音からの半音の数が関数の値となる。

Prev (前音からの音高変化)

前の音からの音高の上下変化を表したものである。

$$Prev = Height(t) - Height(t - 1) \quad (9)$$

Next (次音への音高変化)

次の音への音高の上下変化を表したものである。

$$Next = Height(t + 1) - Height(t) \quad (10)$$

以上まで、各音について物理特徴量 5 個、楽譜情報 5 個が得られた。これらからなる 10 次元のベクトルを、音の一つ一つが持つ音情報ベクトルとする。表 3 に、これをまとめた。

表 3 音情報ベクトル

分類	番号	特徴名称	説明
楽譜情報	1	<i>Bar</i>	小節の頭の音であるか否か
	2	<i>Beat</i>	拍頭の音であるか否か
	3	<i>Height</i>	音符の高さ (音番号)
	4	<i>Prev</i>	前の音からの音高の変化
	5	<i>Next</i>	次の音への音高の変化
物理特徴	6	<i>Length</i>	音の長さの相対値
	7	<i>Pitch</i>	音程の平均律からのずれ
	8	<i>PitchWidth</i>	前音との音程の間隔
	9	<i>Power</i>	音の強さの相対値
	10	<i>Stand</i>	音の立ち上がりの強さ

2.4 演奏特徴

前節までで、10要素からなる音情報ベクトルを求めた。次に、多数の音情報ベクトルから、一つの演奏特徴ベクトルを抽出する。

抽出方法

一演奏に含まれる音の数を N とする。この時得られる、 N 個の音情報ベクトル中の各要素の平均と分散、また各要素間の相関を演奏特徴とする。具体的に、演奏特徴ベクトルの要素を表 4 に列挙した。

各特徴の説明

演奏特徴ベクトルの各特徴が、具体的にどのような意味を持っているのかを、いくつか例にとって説明する。

- 1.*Mean*(7):音程の平均。この値が大きいほど音程が高めの演奏である。
 - 5.*Var*(6):音の長さの分散。この値が大きいほど、ゆらぎの大きい演奏である。
 - 9.*Var*(9):音の強さの分散。この値が大きいほど、起伏の大きい演奏である。
- このように、音のみから得られる特徴に加え、
- 15.*Cor*(1,9):小節頭と音の強さの相関。この値が大きいほど、小節頭に重きを置いた演奏である。
 - 28.*Cor*(3,10):音高と音の立ち上がりの相関。音が高いほど鋭い演奏となることを示す。
- など、楽譜情報を反映した特徴もとれている。

3. 実験

本章では、実際に市販の CD から演奏者の演奏特徴ベクトルを抽出し、各奏者の個性を学習する。さらに、未学習の演奏の奏者識別実験を行う。また、各演奏者の定性的な傾向についても論じる。

実験の基本的な手順を以下に示す。

学習サンプル演奏一回あたりに含まれる音数を N_{learn} 、テストサンプル演奏一回あたりに含まれる音数を N_{test} とする。

表 4 演奏特徴ベクトル

番号	内容	番号	内容	番号	内容
1	<i>Mean</i> (7)	21	<i>Cor</i> (3,8)	41	<i>Cor</i> (8,9)
2	<i>Mean</i> (8)	22	<i>Cor</i> (3,9)	42	<i>Cor</i> (8,10)
3	<i>Mean</i> (10)	23	<i>Cor</i> (3,10)	43	<i>Cor</i> (9,10)
4	<i>Var</i> (6)	24	<i>Cor</i> (4,6)		
5	<i>Var</i> (7)	25	<i>Cor</i> (4,7)		
6	<i>Var</i> (8)	26	<i>Cor</i> (4,8)		
7	<i>Var</i> (9)	27	<i>Cor</i> (4,9)		
8	<i>Var</i> (10)	28	<i>Cor</i> (4,10)		
9	<i>Cor</i> (1,6)	29	<i>Cor</i> (5,6)		
10	<i>Cor</i> (1,7)	30	<i>Cor</i> (5,7)		
11	<i>Cor</i> (1,8)	31	<i>Cor</i> (5,8)		
12	<i>Cor</i> (1,9)	32	<i>Cor</i> (5,9)		
13	<i>Cor</i> (1,10)	33	<i>Cor</i> (5,10)		
14	<i>Cor</i> (2,6)	34	<i>Cor</i> (6,7)		
15	<i>Cor</i> (2,7)	35	<i>Cor</i> (6,8)		
16	<i>Cor</i> (2,8)	36	<i>Cor</i> (6,9)		
17	<i>Cor</i> (2,9)	37	<i>Cor</i> (6,10)		
18	<i>Cor</i> (2,10)	38	<i>Cor</i> (7,8)		
19	<i>Cor</i> (3,6)	39	<i>Cor</i> (7,9)		
20	<i>Cor</i> (3,7)	40	<i>Cor</i> (7,10)		

Mean は平均,*Var* は分散,*Cor* は相関係数を表す。

関数中の数字は、音情報ベクトルの要素番号である。(表 3 参照)

(1) 学習データ演奏から、 N_{learn} 音を一演奏とする演奏特徴ベクトルをランダムに多数生成する。これが学習サンプルとなる。

(2) 線形判別分析により、各奏者の分離を行う。

(3) 未学習演奏中の任意の箇所から連続する N_{test} 個の音を切り出し、これをテストサンプルとする。

(4) テストサンプルの演奏特徴ベクトルを 2 で得た判別空間に射影し、学習サンプルのクラスタとの距離を規準とし識別を行う。

演奏特徴ベクトルは、演奏中に含まれる音の統計的特徴なので、含む音の数が多いほど統計的に安定し、少ないほどバラつきが大きくなるはずである。したがって、般化能力の面から考えると、 N_{learn} はある程度小さくすることが好ましいが、あまりに小さすぎると今度は意味のない学習サンプルばかりとなり学習効率が落ちると思われる。また、 N_{test} は大きいほど認識がよくなると思われる。

3.1 奏者の個性の学習

学習をバルティータ第 1 番で行う。この曲は、第 1 楽章から第 4 楽章まで全て同じ調 (h-moll) である。学習には、第 1 楽章、第 2 楽章、第 4 楽章を用いた。

各奏者の演奏特徴ベクトルを判別平面空間上に射影したものを図 9 に示す。横軸が第一判別軸、縦軸が第二判別軸である。

判別係数の値を見ることで、個性がどのような特徴に表れるかを評価することができる。特に判別係数の絶対値が大きい特徴を、表5に示した。

表5 各判別軸において重要な特徴

軸	特徴	内容	判別係数
第一判別軸	1. Mean(7)	音程の平均	-2.28
	2. Mean(8)	音程間隔の平均	-0.92
	6. Var(8)	音程間隔の分散	-0.90
	16. Cor(2,8)	拍頭と音程間隔の相関	-1.11
第二判別軸	7. Var(10)	音強の分散	-0.92
	22. Cor(3,10)	音高と音強さの相関	-1.02
	26. Cor(4,8)	音高変化と音程間隔の相関	0.89
	38. Cor(7,8)	音程と音程間隔の相関	0.88

第一、第二軸とも音程に関わる特徴が重要となっているが、第二軸は音強に関わる特徴も表れている。これらの特徴と、判別空間上での位置から奏者の傾向が議論できる。以下、いくつか例にとって奏者の傾向を示す。

Player5は第一判別軸の値が大きい。音程が低く、音程の間隔が狭い。音程の変化が少ない演奏であると言える。

Player3は第二判別軸の値が大きい。音の高さに関わらず音強の変化が少ない、淡々とした演奏であると言える。

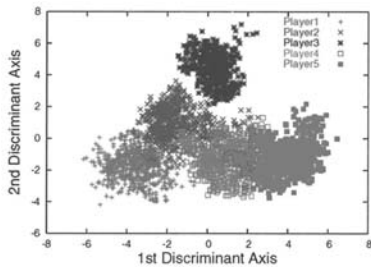


図5 各奏者の判別空間での位置

3.2 同じ調の未学習曲で奏者識別を行う場合

未学習サンプルをパーティータ第1番第3楽章の演奏からとった。これを、学習で得られた判別空間に射影した結果図6のようになった。未学習サンプルが、学習サンプルのクラスタに近い位置にプロットされていることが分かる。このことから、異なる曲であっても同じ調であれば、演奏者の演奏特徴はそれほど変化しないことが分かる。

次に、 N_{learn} を100に固定し、 N_{test} を変えて認識実験を行った結果を表6にまとめた。未学習サンプルの音の数 N_{test} が増えるほど認識がよくなっていることが分かる。

3.3 異なる調の未学習曲で奏者識別を行う場合

今度は、未学習サンプルとして、学習時と異なる調の楽曲を認識実験を行う。未学習サンプルとしてソナタ1番の第4楽章(g-moll)を用いる。各奏者の判別空間上での位置を図7に示した。

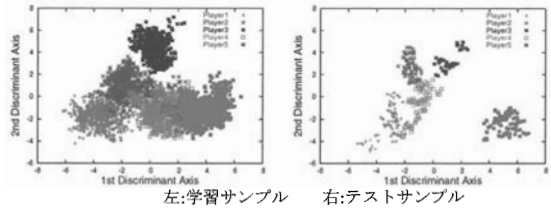


図6 各奏者の判別空間での位置(同じ調の曲)

表6 同じ調の未学習曲における奏者認識率

N_{test}	50	100	150
Player1	70.5%	94.3%	100%
Player2	65.6%	79.0%	92.1%
Player3	66.1%	95.9%	100%
Player4	50.5%	69.0%	90.9%
Player5	67.3%	76.1%	87.2%
Average	64.0%	82.9%	94.0%

N_{learn} は100に固定

これから分かるように、同じ調での認識に比べ、各奏者のクラスタが移動してしまっている。特に、Player1, Player3はクラスタの移動が著しい。

次に、 N_{learn} を100に固定し未学習演奏の識別実験を行った結果を表7にまとめた。クラスタが移動している奏者ほど、認識率が悪いことが分かる。

これらの結果より、異なる調の曲では、奏者によって演奏特徴が大きく変化する場合があることが分かった。

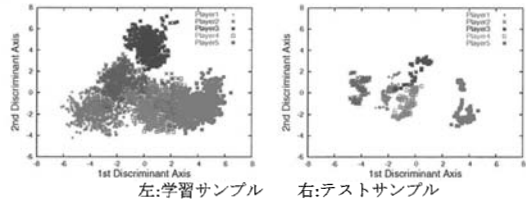


図7 各奏者の判別空間での位置(異なる調の曲)

表7 異なる調の未学習曲における奏者認識率

N_{test}	100	200	400
Player1	31.3%	37.3%	31.6%
Player2	85.8%	84.0%	85.6%
Player3	45.4%	64.3%	74.6%
Player4	85.6%	91.8%	100%
Player5	90.6%	96.9%	100%
Average	67.7%	74.9%	78.4%

N_{learn} は100に固定

3.4 異なる録音環境の演奏を認識する場合

実演奏音は音響特性の影響を強く受ける。このため、本手法で演奏者の個性として抽出した特徴には音響特性の特徴が多く含まれてしまっている可能性がある。この点を検討するため、異なる録音環境の演奏の識別実験を行う。3人の奏者について、同じ曲を収録した2枚のCDを用意し、片方のCDで学習を行い、他方のCD中のサンプルを識別する。使用した曲は、バルティータ第1番の第1楽章、第2楽章、第3楽章、第4楽章である。ただし、各奏者それぞれについて、2枚のCDは録音した時期に10年から20年ほどの隔りがある。このため、音響特性だけでなく、奏者自身のスタイルの変化も問題となる。単純な判別分析ではこれらの要素に強く判別軸がフィットしてしまうため、何らかの方法で学習を適度に抑える必要がある。

正則化項を加えた学習

判別分析の際、学習データのクラス内共分散行列に、適当な大きさの自然行列（正則化項）を加えることにより、オーバーフィッティングを防ぐことができる。これは、学習サンプルに人工的なノイズを加えることと等価である。正則化項を加え学習を適度に抑えることにより、経年変化によらない奏者の大まかな個性を学習することを試みた。

この方法で認識実験を行った結果は次のようになった。図8に各奏者の判別空間上での各奏者の位置、表8に認識率を示した。

これらの結果から、異なる音響特性下の演奏でも認識が可能であることが示された。

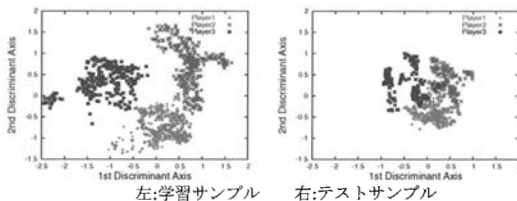


図8 各奏者の判別空間での位置（正則化項あり）

表8 異なる環境下の奏者認識率

N_{test}	200	300	400
Player1	76.8%	83.7%	86.8%
Player2	93.8%	96.6%	100%
Player3	75.9%	81.4%	91.4%
Average	82.2%	87.2%	92.7%

N_{learn} は 150 に固定

4. 結論と展望

4.1 結論

本研究では、実演奏音から演奏者の個性を学習し識別を行う手法を提案した。

線形判別分析により演奏者の演奏特徴ベクトルを分析し、次のような結果を得た。

- 各奏者が判別空間上で分離された。
- 判別空間上の位置から各奏者の個性が論じられた。
- 未学習曲の演奏曲の奏者認識実験では、同じ調の未学習曲では認識率 94.0%、異なる調の未学習曲では認識率 78.4%となり、特に同じ調の場合において良好な結果を示した。また、異なる曲では、奏者によっては演奏が大きく変化する場合があることが分かった。
- 学習を適度に抑えることで、異なる音響特性下や年代の演奏の奏者認識実験でも、識別率 92.7%と良好な結果を得た。

4.2 課題と今後の展開

本研究の手法について、いくつか解決すべき課題を挙げる。

本研究の手法では、信号解析上の問題から扱える曲は限られており、今回用いたような独奏曲で機械的な構造の曲のみである。しかし、演奏家は、曲の作曲された時代や地域、作曲者によっても演奏スタイルを変化させるため、より普遍的な個性を抽出するためには、さまざまな曲から学習を行う必要がある。このため、いろいろなタイプの曲に対応できるよう、信号解析段階からアルゴリズムを改良していく必要がある。

また、実験結果より、本手法は音響特性の影響についてもある程度頑健であることが示されたが、それでも音響特性の影響から逃れることは出来ない。演奏と音響特性の関係を慎重に考察し、より頑健な特徴の抽出手法を考える必要がある。

謝 辞

本研究は、東京大学工学部機械情報工学科知能情報システム研究室において卒業論文として行われたものである。御議論頂いた國吉康夫教授始め研究室の皆様に深く感謝いたします。

文 献

- [1] J.-J. Aucouturier and F. Pachet: Music similarity measures: What's the use?, *Proc. ISMIR 2002*, pp.133-139 (2002).
- [2] M. Goto: A chorus-section detecting method for musical audio signals, *Proc. ICASSP 2003*, V-437-440, (2003).
- [3] 橋田光代, 野池賢二, 片寄晴弘: 声部協調型音楽解釈モデルに基づく演奏の表情付け, 情報処理学会研究報告, 2005-MUS-59, pp.61-66, (2005).
- [4] M. Nishijima and K. Watanabe: Interactive music composer based on neural networks, *Proc. ICMPCC*, pp. 53-56, (2004).
- [5] 浜中雅俊, 後藤真孝, 麻生英樹, 大津展之: 演奏者の振舞いを統計的に学習するジャムセッションシステム, 情報処理学会論文誌, Vol.45, No.3, pp.698-709 (2004).