楽器音を対象とした音源同定：音高による音色変化を考慮する識別手法の検討

北原 鉄朗[†]　後藤 真孝[‡]　奥乃 博[†]

[†] 京都大学大学院情報学研究科知能情報学専攻
[‡] 科学技術振興事業団さきがけ研究21「情報と知」領域 ／ 産業技術総合研究所

kitahara@kuis.kyoto-u.ac.jp　m.goto@aist.go.jp　okuno@i.kyoto-u.ac.jp

あらまし 本稿では，音源同定のための音高による音色変化の表現方法として，F0 依存多次元正規分布を提案する．我々は，これまでに音高による音色変化を適切に扱うことの必要性を指摘し，特徴変動を基本周波数の関数として近似する手法を提案した．しかし，識別関数は木下らの類似度に限られ，他の識別手法に応用することはできなかった．本稿では，より一般的に音高による音色変化をとらえるため，音色空間上で各クラスのパターンが，平均が音高によって変化する多次元正規分布に従うと仮定する．そして，この分布のパラメータの推定法を提案し，識別関数をベイズ決定規則から定式化する．提案手法を実装・実験した結果，通常の多次元正規分布を仮定した場合に比べ，個々の楽器レベルで平均 16.48%，カテゴリーレベルで平均 20.67%の認識誤りを削減することができた．

Musical Instrument Identification: A Classifier Considering Pitch-dependent Characteristics of Timbre

Tetsuro Kitahara[†]　Masataka Goto[‡]　Hiroshi G. Okuno[†]

[†] Dept. of Intelligence Science and Technology,
Graduate School of Infomatics, Kyoto University
[‡] "Information and Human Activity", PRESTO, JST ／
National Institute of Advanced Industrial Science and Technology

Abstract This paper describes a pitch-dependent timbre representation for musical instrument identification, *F0-dependent multivariate normal distribution*. In our previous paper, we proposed a method of curve-fitting the pitch-dependent characteristics and constructing timbre classifiers as a function of fundamental frequency. In this paper, we extended the method to estimate paramaters of the F0-dependent multivariate normal distribution and to define a discriminant function for this distribution according to the Bayes decision theory. Experimental results showed that the proposed method improved the performance of musical instrument identification.

1. はじめに

楽器音の同定がパターン認識の研究対象として広く扱われるようになったのは，音声や文字などより遅く，1990 年代に入ってからのことである[1)~8)]．そのため，音声認識や文字認識に比べ，特徴抽出や識別手法に関して得られている知見は少ない．また，楽器音響学の分野では古くからさまざまな分析が行われてきた[9)~12)]が，音源同定の工学的モデルの実現には至っていない．

我々は，音源同定を**音楽情景分析**[3)]（音楽音響信号を対象とした計算機による聴覚的情景分析[13)]）の重要な要素技術の 1 つと位置づけ，研究を進めている．通常，聴覚的情景分析という場合には混合音を扱う場合が多い．しかし，音源同定は単音でも難しい問題であり，人間の単音に対する音源同定能力は，音楽経験者であっても 14 種類の楽器に対して 45.9%という報告もある[6)]．そこで，我々は (1) 単音の音源同定，(2) 混合音の音源同定，という 2 段階のアプローチをとって

研究を進めている．本稿では，第 1 段階の単音の音源同定について報告する．

我々は，これまでに音高による音色変化に着目した識別手法について研究してきた[14]．楽器音における特徴変動要因としては，音高，音の強さ，楽器の個体差，奏法などが考えられるが，これらのうち音高は物理量（基本周波数）として抽出可能である．そこで，音高による特徴変動を基本周波数の関数として近似して識別器を設計する手法を提案した[14]．このように，特徴変動をその要因となる物理量の関数としてとらえるアプローチは，従来の研究[1)~8)]では行われてこなかった．しかし，文献 14) の手法は，(1) 識別関数として木下らの類似度[5)]を基としており，他の識別関数に応用することができない，(2) 各特徴量を音高による変化の仕方に基づいて手動で分類しなければならない，という問題があった．また，使用データベースの楽器数が少なく，大規模なデータベースで有効性を確認するには至らなかった．

本稿では，音高による音色変化をより一般的に表現する方法として，音高によって平均が変化する多次元正規分布（**F0 依存多次元正規分布**）を提案する．そして，各クラスのパターンが音色空間（楽器音に関する特徴空間）上で，この F0 依存多次元正規分布に従うと仮定し，ベイズ決定規則に基づいて識別関数を定式化する．なお，音高による音色変化の関数近似では，大規模な楽器音データベース[15)]の一部を使用してデータ量を増やすことで，文献 14) のように特徴量を分類せずに，より高次の曲線で近似する．

以下，まず **2.** で F0 依存多次元正規分布を提案し，この分布を仮定した場合の識別関数をベイズ決定規則に基づいて定式化する．次に，**3.** で提案手法の処理の流れを述べ，**4.** で評価実験について述べる．さらに，**5.** でベイズ決定規則と k-NN 法を比較し，最後に **6.** でまとめをする．

2. F0 依存多次元正規分布

本稿で述べる音源同定方式では，各楽器名がラベルづけられた楽器音の音響信号のデータベース（個々の楽器音データを学習パターンと呼ぶ）に基づいて音源同定を行う．各楽器の学習パターンが多次元正規分布に従うと仮定し，多次元正規分布のパラメータを推定して各楽器の事後確率を計算する．そして，事後確率の最も高い楽器名を同定結果として出力する．ただし，学習パターンは，以下の理由により音高に依存する：

(1) 音高が低くなれば，発音体は大きくなる．発音体の質量が大きくなると慣性も大きくなり，発音の立ち上がりや減衰に，より多くの時間を要する[10)]．

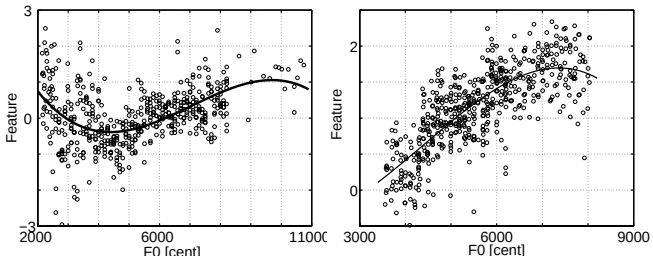


図 1 代表値関数（太字）の例．左の図は線形近似では精度が不十分な例（ピアノの第 4 軸）で，右の図は音高による音色変化が特に顕著な例（チェロの第 1 軸）である．

(2) 音高が高くなると振動損失が大きくなるため，高次の高調波は発生されにくくなる[10)]．

(3) 一部の楽器では音高により発音体が異なり，各発音体は異なる材質からできている．

この問題に対する 1 つの解決法は，各楽器の学習パターンが音高ごとに異なる多次元正規分布に従うと仮定し，対応する音高の学習パターンのみを使って各音高における分布のパラメータを推定することである．しかし，分布のパラメータ推定には多くの学習パターンを必要とし，音高ごとに多くの学習パターンを用意するのは非現実的である．

本稿では，各楽器の学習パターンが，平均のみが音高ごとに異なる多次元正規分布に従うと仮定する．すなわち，音高ごとに用意された多次元正規分布は，共分散行列がすべて等しいものと仮定する．このように仮定すると，共分散行列は，音高以外の要因による音色変化を表していると考えることができ，すべての音高の学習データを用いて算出することができる．本稿では，このような音高によって平均が変化する多次元正規分布を **F0 依存多次元正規分布**と呼ぶ．以下，F0 依存多次元正規分布のパラメータ推定法を述べる．

2.1 代表値関数

音高によって変化する分布の平均を，最小二乗法による関数近似で推定する（**図 1**）．文献 14) では，特徴量を音高による変化の仕方にしたがって手動で分類し，区分的線形関数によって近似していた．データ数が多ければより高次の関数で近似することが可能であり，より高次の関数を用いることができれば，表現できる曲線の自由度が上がるため，変化の仕方にしたがって特徴量を分類する必要はなくなる．そこで，本稿では区分的線形関数ではなく 3 次関数を用いる．この近似曲線を**代表値関数**と呼び，$\boldsymbol{\mu}_i(f)$ と書く（i：クラス番号）．

2.2 F0 正規化共分散行列

F0 依存多次元正規分布における共分散行列の算出法を述べる．共分散行列は，音高以外の要因による音色変化を表していると考えることができる．そこで，音

色空間を代表値関数で正規化することで音高による音色変化を除去してから，共分散行列を求める．この方法で求めた共分散行列を**F0 正規化共分散行列**と呼び，Σ_iと書く．

2.3 ベイズ決定規則による識別

ベイズ決定規則に基づいて識別関数を定式化する．各クラスω_iのパターンが確率密度関数 $p(\boldsymbol{x}|\omega_i)$ に従うと仮定し，パターン$\boldsymbol{x}$が入力されたときの識別関数を

$$g_i(\boldsymbol{x}) := \log p(\boldsymbol{x}|\omega_i) + \log p(\omega_i) \qquad (1)$$

と定義する．ここで，$p(\omega_i)$ はクラスω_iの事前確率である．そして，$g_i(\boldsymbol{x})$ が最大になる i を求める．すなわち，$k := \mathrm{argmax}_i\, g_i(\boldsymbol{x})$ とすると，クラスω_kを同定結果とする．

本稿では，各クラスのパターンが，前節までに述べた F0 依存多次元正規分布に従うと仮定する．このとき，入力パターン$\boldsymbol{x}$の基本周波数を f とすると，クラスω_iの確率密度関数 $p(\boldsymbol{x}|\omega_i; f)$ は，

$$p(\boldsymbol{x}|\omega_i; f) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}D^2(\boldsymbol{x}, \boldsymbol{\mu}_i(f))\right\}$$

で与えられる．ここで，d は音色空間の次元数，D はマハラノビス距離で

$$D^2(\boldsymbol{x}, \boldsymbol{\mu}_i(f)) := (\boldsymbol{x} - \boldsymbol{\mu}_i(f))'\Sigma_i^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_i(f))$$

である（$'$は転置）．この式を式 (1) に代入すると次の識別関数 g が得られる：

$$\begin{aligned} g_i(\boldsymbol{x}; f) &:= \log p(\boldsymbol{x}|\omega_i; f) + \log p(\omega_i; f) \\ &= -\frac{1}{2}D^2(\boldsymbol{x}, \boldsymbol{\mu}_i(f)) - \frac{1}{2}\log|\Sigma_i| \\ &\qquad -\frac{d}{2}\log 2\pi + \log p(\omega_i; f). \end{aligned}$$

$p(\omega_i; f)$ は入力パターンの基本周波数が f のときのクラスω_iの事前確率である．そして，$k := \mathrm{argmax}_i\, g_i(\boldsymbol{x}; f)$ とおいて，クラスω_kを同定結果として出力する．

ここで，事前確率 $p(\omega_i; f)$ の求め方を述べる．これは，入力パターンの基本周波数が f であるときに，各クラスの音域に f が含まれているかを表す．すなわち，クラスω_iの音域を R_iとすると，クラスω_iの事前確率を

$$p(\omega_i; f) := \begin{cases} 1/m & (\text{if } f \in R_i) \\ 0 & (\text{if } f \notin R_i) \end{cases}$$

と定義する．ここで，m は各クラスの事前確率の合計を 1 にするための正規化定数である．

3. 処理の流れ

本章では，提案手法の処理の流れを述べる．まず，前処理としてスペクトログラムを作成し，調波構造を推定する．次に，特徴抽出を行う．特徴抽出は，後で特徴空間の変形（次元圧縮）をすることを前提に，識別に有効と期待できる特徴を数多く抽出する．その後，主成分分析・線形判別分析により次元圧縮を行う．そして，圧縮された音色空間上で各クラスのパターンが F0 依存多次元正規分布に従うと仮定し，ベイズ識別規則を用いて楽器名を同定する．

3.1 調波構造の推定

まず，短時間フーリエ変換を用いてスペクトログラムを作成し（ハニング窓使用，窓幅：4096 点，時間分解能：10ms），各時刻において，パワースペクトログラムの周波数方向の導関数の零交差からピーク抽出を行う．ピーク位置推定には，複素スペクトル内挿法[16]をハニング窓用に拡張したもの[17]を使用し，抽出されたピークから調波構造（30 次まで）を抽出する．なお，基本周波数に関しては，音名を人手で与え，その音名に対応する周波数（平均律で算出）の近傍（200cent いない）に存在するピークの周波数とする．また，周波数とパワーは，ともに対数で表し，正規化は特に行わない．

3.2 特 徴 抽 出

次に示す 129 個の特徴量を抽出する．これらは，先行研究[3),6)]や楽器音響学・楽器物理学などの知見[9)~12)]を参考にしながら決定した．

(1) スペクトルに関する特徴

ここでは，主に音の甲高さなどスペクトルの定常的な特徴を抽出する．そのため，各高調波成分の周波数やパワー値は，その時間方向の中央値を用いる．具体的には次に示す 40 個の特徴量を抽出する：

1 周波数重心（各高調波成分のパワー値を重みとする周波数の重みつき平均），

2 全高調波成分のパワー値の合計に対する基音成分のパワー値の割合，

3〜30 全高調波成分のパワー値の合計に対する基音から i 次までの高調波成分のパワー値の合計の割合 $(i = 2, 3, \cdots, 29)$，

31 奇数次の高調波成分（基音含む）と偶数次の高調波成分とのパワー値の合計の比，

32〜40 音がなり続けている時間（周波数成分全体のパワーがしきい値を越えている時間）に対して，その高調波成分の鳴り続けている時間（パワー値が同じしきい値を越えている時間）が p%である高調波成分の個数 $(p = 10, 20, \cdots, 90)$．

(2) パワーの時間変化に関する特徴

ここでは，パワーの時間変化に関する特徴を抽出する．以下の特徴量41で，大局的な音量変化（通常，音が減衰するか持続するか）を表し，特徴量42〜75で，より細かな音量変化を表す．

41 パワー包絡線の線形最小二乗法による近似直線の傾き，

42〜58 発音開始直後 t 秒間のパワー包絡線の微分

係数の中央値 ($t = 0.15, 0.20, \cdots, 0.95$),

59〜75 最大パワー値と，発音開始から t 秒後のときのパワー値の比 ($t = 0.15, 0.20, \cdots, 0.95$).

(3) **各種変調の振幅と振動数**

以下の変調の振幅と振動数を抽出する．ここで，各種変動の振動数は導関数の零交差点数から，振幅は，十分に平滑化された信号と元の信号との差に対する四分位幅（上位25%と下位25%の値を無視したときの最大値と最小値の差）からそれぞれ算出する．平滑化には，Savitzky と Golay の 2 次多項式適合による平滑化法[18]を使用する．

76 振幅変調の振幅,
77 振幅変調の振動数,
78 周波数変調の振幅,
79 周波数変調の振動数,
80 周波数重心の時間変化の振幅,
81 周波数重心の時間変化の振動数,
82〜94 k 次のメル周波数ケプストラム係数(MFCC)の時間変化の振幅 ($k = 1, 2, \cdots, 13$),
95〜107 k 次の MFCC の時間変化の振動数 ($k = 1, 2, \cdots, 13$).

(4) **発音開始直後のピーク尖度に関する特徴**

発音開始直後 150ms 間において，各高調波成分のピーク周辺にどの程度非調波成分があるかを，各高調波成分のピークの尖度から抽出する．まず，発音開始時刻から 150ms までの各時刻のパワースペクトルから，基音から 11 次倍音までの各高調波成分のピーク付近（ピークの周波数を f[Hz] とすると，$0.75f$[Hz] から $1.5f$[Hz] までの範囲）の部分を切り出す．そして，切り出された各ピーク付近がどの程度とんがっているかを 4 次モーメントから算出する．このとき，非高調波成分が多く含まれていれば，高調波成分のピークが非高調波成分に埋もれる形となるため，ピークの尖度は低くなり，逆に，非高調波成分があまり含まれていなければ，ピークの尖度は高くなる．そこで，基音から 11 次倍音までの各高調波成分に対する各時刻のピーク尖度の時間方向の平均値（特徴量番号108〜118）と，時間変化の振幅（特徴量番号119〜129）をそれぞれ抽出する．

3.3 主成分分析・線形判別分析による次元圧縮

抽出された特徴量を平均が 0，分散が 1 になるように正規化し，主成分分析により次元を圧縮する．累積寄与率 99%で，特徴空間は 129 次元から 79 次元に圧縮される．

次に，線形判別分析によりさらに次元を圧縮する．本稿の実験では 19 種類の楽器を扱うので，特徴空間は 18 次元に圧縮される．線形判別分析は，クラス内分散・クラス間分散比を最大にする部分空間を求める手法で，識別を考慮した次元圧縮法である．そのため，主成分分析のみで同次元に圧縮するのに比べて高性能になると予測される．このことは，後述の実験で確認する．

表 1 使用した楽器音データベースの内訳

楽器番号	楽器名（楽器記号）	楽器個体	音域	強さ	奏法	データ数
01	ピアノ (PF)	3	A0–C8			508
09	クラシックギター (CG)	3	E2–E5			696
10	ウクレレ (UK)	3	F3–A5			295
11	アコースティックギター (AG)	3	E2–E5	そ		666
15	バイオリン (VN)	3	G3–E7	れ		528
16	ビオラ (VL)	3	C3–F6	ぞ		472
17	チェロ (VC)	3	C2–F5	れ	通	558
21	トランペット (TR)	2	E3–A♯6	強	常	151
22	トロンボーン (TB)	3	A♯1–F♯5	・	の	262
25	ソプラノサックス (SS)	3	G♯3–E6	中	奏	169
26	アルトサックス (AS)	3	C♯3–A5	・	法	282
27	テナーサックス (TS)	3	G♯2–E5	弱	の	153
28	バリトンサックス (BS)	3	C2–A4	の	み	215
29	オーボエ (OB)	2	A♯3–G6	3		151
30	ファゴット (FG)	3	A♯1–D♯5	種		312
31	クラリネット (CL)	3	D3–F6	類		263
32	ピッコロ (PC)	3	D5–C8			245
33	フルート (FL)	3	C4–C7			134
34	リコーダー (RC)	3	C4–B6			160

データ数：無音検出による自動切り出しによって切り出された単音の個数.

3.4 識　　別

2. で述べたように，主成分分析・線形判別分析によって圧縮された 18 次元の特徴空間上で，各クラスのパターンが F0 依存多次元正規分布に従うと仮定し，ベイズ決定規則を用いて楽器名を同定する．

4. 評 価 実 験

提案手法の有効性を確認するため，評価実験を行う．

4.1 実 験 方 法

実楽器の単音データベースとして，RWC 研究用音楽データベースの楽器音データベース RWC-MDB-I-2001[15]を使用した．これは，50 種類の実楽器の単独発音を半音ごとに収録（サンプリング周波数：44.1kHz, 16 ビットリニア量子化，モノラル）したもので，各楽器音には，原則 3 種類の楽器個体，3 種類の音の強さ，複数の奏法が含まれている．

このデータベースのうち，オーケストラで一般的に使用される楽器から，打楽器，ノイズが大きいもの，音高が著しく不安定なものなどを除いた 19 種類の楽器を使用した．使用したデータ（総数：6247 個）の内訳を**表 1** に示す．表 1 のデータ全体を無作為に 10 等分し，クロスバリデーションを行って認識率を求める．すなわち，10 個のグループそれぞれに対して，そのグループ以外のデータで学習してそのグループのデータ

表 2　19 楽器の分類

カテゴリー	属する楽器
ピアノ	ピアノ
ギター	クラシックギター，ウクレレ， アコースティックギター
弦楽器	バイオリン，ビオラ，チェロ
金管楽器	トランペット，トロンボーン
サックス	ソプラノサックス，アルトサックス， テナーサックス，バリトンサックス
複簧楽器	オーボエ，ファゴット
クラリネット	クラリネット
無簧楽器	ピッコロ，フルート，リコーダ

で評価という実験を繰り返す．以下で述べる認識率は，10 回の平均である．

楽器音を扱う場合，個々の楽器の認識率だけでなく，弦楽器，金管楽器などのカテゴリーレベルの認識率も重要である．なぜなら，実際の応用においてカテゴリーレベルの情報が分かるだけで有用な場面が多いからである．たとえば，ピアノソロ曲を検索する場面では，音楽音響信号に弦楽器（擦弦楽器）や管楽器などが含まれていることが分かれば，それだけで検索対象からはずすことができる．また，フルートとピアノのアンサンブル曲を自動採譜する場面で，個々の楽器名を正しく同定できなくても両者を聞き分けることはできる．

本稿では，カテゴリーレベルの認識率を**表 2** に示す 8 つのカテゴリーを用いて算出する．これは，楽器の発音機構や従来研究[6),8)]に基づいて本研究で定義した．ただし，Eronen は，複簧楽器とクラリネットを 1 つのカテゴリーに，金管楽器とサックスを 1 つのカテゴリーにまとめた 6 カテゴリーを用いている[8)]．ここで，本稿で用いる「カテゴリー」という言葉は，弦楽器，金管楽器など，個々の楽器名（バイオリン，トランペットなど）より一段抽象度の高い概念を表し，音源同定システムが出力する概念としての「クラス」とは区別する．

4.2 実験結果

通常の多次元正規分布を仮定して識別した場合と F0 依存多次元正規分布を仮定して識別した場合の両方の認識率を**表 3** に示す．本稿で提案する F0 依存の処理を導入することで，個々の楽器レベルで，平均の認識率が 75.73%から 79.73%と 4.00%改善され，誤認識が 16.48%削減された．また，カテゴリーレベルでは，平均の認識率は 88.20%から 90.65%と 2.45%改善され，誤認識が 20.67%削減された．

これらの結果が有意であることを以下に検定で示す．帰無仮説 H_0 と対立仮説 H_1 をそれぞれ H_0：（F0 依存多次元正規分布の場合の認識率）$\leq$（通常の多次元正規分布の場合の認識率），H_1：（F0 依存多次元正規分布の場合の認識率）$>$（通常の多次元正規分布の場合の認識率）として，t 検定（片側検定）を行う．各楽器

表 3　実験結果（通常の多次元正規分布の場合の認識率と F0 依存多次元正規分布の場合の認識率）

楽器	個々の楽器レベル			カテゴリーレベル		
記号	Normal	F0-dpt	差	Normal	F0-dpt	差
PF	74.21%	83.27%	+9.06%	74.21%	83.27%	+9.06%
CG	90.23%	90.23%	±0.00%	97.27%	97.13%	−0.14%
UK	97.97%	97.97%	±0.00%	97.97%	98.31%	+0.34%
AG	81.23%	83.93%	+2.70%	94.89%	95.65%	+0.76%
VN	69.70%	73.67%	+3.97%	98.86%	99.05%	+0.19%
VL	73.94%	76.27%	+2.33%	93.22%	94.92%	+1.70%
VC	73.48%	78.67%	+5.19%	95.16%	96.24%	+1.08%
TR	73.51%	82.12%	+8.61%	76.82%	85.43%	+8.61%
TB	76.72%	84.35%	+7.63%	85.50%	89.69%	+4.19%
SS	56.80%	65.89%	+9.09%	73.96%	80.47%	+6.51%
AS	41.49%	47.87%	+6.38%	73.76%	77.66%	+3.90%
TS	64.71%	66.01%	+1.30%	90.20%	92.16%	+1.96%
BS	66.05%	73.95%	+7.90%	81.40%	86.05%	+4.65%
OB	71.52%	72.19%	+0.67%	75.50%	74.83%	−0.67%
FG	59.61%	68.59%	+8.98%	64.74%	71.15%	+6.41%
CL	90.69%	92.07%	+1.38%	90.69%	92.07%	+1.38%
PC	77.56%	81.63%	+4.07%	89.39%	90.20%	+0.81%
FL	81.34%	85.07%	+3.73%	82.09%	85.82%	+3.73%
RC	91.88%	91.25%	−0.63%	92.50%	91.25%	−1.25%
平均	75.73%	79.73%	+4.00%	88.20%	90.65%	+2.45%

Normal: 通常の多次元正規分布を仮定した場合
F0-dpt: F0 依存多次元正規分布を仮定した場合（提案手法）

における両手法の認識率の差を d_i $(i = 1, \cdots, n)$ とすると，検定統計量 t_0 は，

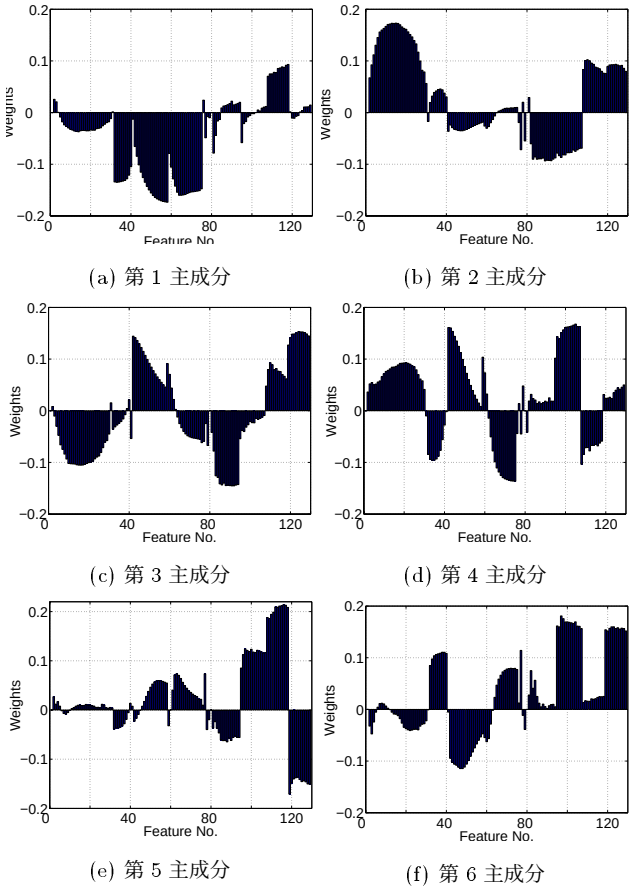
$$t_0 := \frac{|\bar{d}|}{\sqrt{\sum_i (d_i - \bar{d})^2 \Big/ n(n-1)}}$$

で与えられる．ここで，$\bar{d}$ は $d_1, \cdots, d_n$ の平均値である．この統計検定量は，個々の楽器レベルとカテゴリーレベルでそれぞれ 5.4781, 3.9482 で，ともに有意水準 0.05%（棄却域：$(3.9217, \infty)$）で有意である．

4.3 主成分分析に関する考察

各主成分の重みの一部を**図 2** に示す．図 2(a), (b) から，第 1 主成分は高調波成分の個数（32〜40）とパワーの時間変化（41〜75）を，第 2 主成分は高調波成分に関する定常的特徴（2〜30）を総合的に表していると考えられる．従来より，音色を規定する要因としてスペクトルに関する定常的特徴とパワーの時間変化に関する特徴が重要であると言われており☆，このことを裏付ける結果になったといえる．その他，第 3 主成分は MFCC や発音開始直後のピーク尖度の時間変化の振幅（82〜94，119〜129），第 4 主成分は MFCC の時間変化の振動数（95〜107），第 5 主成分は MFCC と発音開始直後のピーク尖度（95〜129），第 6 主成分は MFCC の時間変化の振動数（95〜107）と発音開始直後のピーク尖度の時間変化の振動数（119〜129）な

☆ たとえば，古典的な音合成方式では，周波数エンベロープオシレータにより定常的なスペクトルを制御し，パワーエンベロープオシレータによりパワーの時間変化を作り出している[19)]．



(a) 第 1 主成分　(b) 第 2 主成分

(c) 第 3 主成分　(d) 第 4 主成分

(e) 第 5 主成分　(f) 第 6 主成分

図 2　主成分分析による各主成分の重み値

どを総合的に表していることが分かる．

ここで，発音開始直後のピーク尖度に関する特徴量（95〜129）が，第 3 主成分，第 5 主成分，第 6 主成分と多くの主成分に現れている．この特徴量は，高調波成分周辺の非調波成分の多さをモデル化したものである．楽器音の非調波成分のモデル化は，従来からその必要性が認識されながらも☆[12),20)]，ほとんど考慮されてこなかった．これらの特徴量が多くの主成分に現れたことは，音楽情景分析において非調波成分を適切に扱う必要があることを示唆している．

4.4　線形判別分析に関する考察

主成分分析と線形判別分析によって特徴空間変形を行った際の重み（基底ベクトル）の一部を**表 4** に示す．表から以下の考察が得られる：

(1)　**スペクトルに関する特徴量について**

第 9 軸に 2（全高調波成分のパワー値の合計に対する基本成分のパワー値の割合）の重みが 0.3586 と高く現れた他は，32〜40（音がなり続けている時間に対して，その高調波成分のなり続けている時間が p%以上である高調波成分の個数）の重みが，第 5 軸，第 7 軸，第 9 軸，第 10〜13 軸で高かった（絶対値で 0.2721〜0.4363）．音色を規定する要因としてスペクトルに関する特徴量が重要であることは以前から知られており[21),22)]，上記の結果は，これを裏付ける結果になったといえる．

(2)　**パワーの時間変化・各種変調について**

41（パワー包絡線の線形最小二乗法による近似直線の傾き）の重みが，第 3 軸，第 4 軸で高く（それぞれ $0.5977, -0.2578$），42（発音開始直後 150ms 間のパワー包絡線の微分係数の中央値）の重みが第 10 軸で -0.3200，73〜75（最大パワー値と，発音開始から $t(= 0.85, 0.90, 0.95)$ 秒後のときのパワー値の比）の重みが第 1 軸で 0.2701〜0.3926 であった．また，76〜79（振幅変調，周波数変調の振幅／振動数）は，第 1 軸，第 2 軸，第 4〜6 軸，第 8 軸，第 12 軸，第 14 軸と，多くの軸で大きな重みが現れた（絶対値で 0.2755〜0.4425）．これらから，パワーの時間変化や各種変調などの動的特徴が識別に効果的であるといえる．実際，人間の音色知覚においても，このような動的特徴が重要であることが知られており[12),21),22)]，楽器音の音響信号を逆転再生すると音源同定精度が低下するという実験結果[23)]などもこのことを裏付けている．

(3)　**発音開始直後のピーク尖度に関する特徴について**

108，109（発音開始直後の基音成分／2 次高調波成分のピーク尖度）が，第 3〜8 軸，第 10 軸，第 11 軸と，多くの軸で大きな重みが現れた（絶対値で 0.2607〜

表 4　特徴空間変形の基底ベクトルの一部

	特徴量と重み値
第 1 軸	73 (0.2701), 74 (0.3220), 75 (0.3926), 79 (−0.3204), 81 (0.2559)
第 2 軸	40 (−0.2721), 76 (0.4425), 78 (0.3554), 82 (−0.2771),
第 3 軸	41 (0.5977), 109 (0.2607)
第 4 軸	41 (−0.2578), 79 (−0.2917), 109 (0.2944)
第 5 軸	40 (0.4286), 78 (0.3219), 108 (0.5400)
第 6 軸	76 (−0.2755), 108 (−0.4529)
第 7 軸	40 (0.3974), 108 (−0.4576)
第 8 軸	76 (0.3378), 85 (0.2614), 108 (−0.4541)
第 9 軸	2 (0.3586), 40 (−0.2783), 84 (0.4525)
第 10 軸	40 (0.2887), 42 (−0.3200), 108 (−0.3292), 109 (0.4508)
第 11 軸	32 (0.4363), 36 (−0.2837), 109 (−0.2732)
第 12 軸	39 (0.2794), 78 (0.3174), 81 (0.2704)
第 13 軸	40 (0.3521), 120 (−0.2522)
第 14 軸	76 (−0.3484), 77 (0.4201)

☆ 文献 20) では,「楽器音を特徴づける重要な要素の一つに音の立ち上がり部の非定常波成分があるが，従来のシステムでは，信号処理的な困難もあり，ほとんど考慮されていなかった」と述べられており，また，文献 12) では，非調波成分を「雑音的成分の混在」と称し,「楽器音の “それらしさ” を構成する重要な因子の 1 つ」と述べている．

0.5400）．これは，**4.3** でも述べたように，音楽情景分析において非調波成分を適切に扱う必要があることを示唆している．

4.5 実験結果に関する考察

実験結果について考察する．

(1) ピアノの性能改善が顕著（74.21%から 83.27%，9.06%の改善）である．これは，ピアノの音域が広く，音高による音色変化が顕著に現れるからと考えられる．楽器音響の分野では，ピアノの音色は，

- 低音ほど倍音が豊富である[12]，
- 低音ほど弦が太く，弦の質量が大きくなるため，振幅の時間変化が緩やかになる[10]，
- 低音では 1 本，中音では 2 本，高音では 3 本の弦が 1 つの鍵盤に対して使われており，中音・高音では調律の微妙なずれにより “うなり” が発生する[12]，

ということが知られている．実際，主成分分析・線形判別分析で得られた特徴空間において，第 2 軸，第 4 軸，第 14 軸に音高による特徴変動が顕著にみられた．これらの軸は，41 パワー包絡線の線形最小自乗法による近似直線の傾き，76〜79 振幅変調・周波数変調の振幅／振動数，などの重みが大きく，上記と一致する部分がみられる．

(2) クラシックギター，ウクレレ，リコーダーでは，提案手法の有効性を確認することはできなかった．これは，元々の認識率が 90%を越えており，改善の余地が小さかったからと考えられる．

(3) ギター，弦楽器のカテゴリーレベルの認識率が，他の楽器に比べ高かった（94.92%〜99.05%）．これは，管楽器は種類が多く，いくつかのカテゴリーにまたがって存在するのに対し，ギターや弦楽器は，他のカテゴリーに発音機構の似た楽器が存在しないためと考えられる．

(4) サックスは，カテゴリーレベルの認識率（77.66%〜92.16%）に比べ，個々の楽器レベルの認識率（47.87%〜73.95%）が低かった．これは，サックス内の個々の楽器の音色が非常に似ているためと考えられる．実際，これらは人間でも識別するのが難しく，文献 6) によれば，被験者が聴いた音の楽器名を 27 個の楽器名が書かれたリストから選ぶ，という実験で正しく認識できた人（音楽経験者）は，ソプラノサックスで 7.1%，アルトサックスで 28.6%と少なかった．ただし，この聴取実験では 10 秒程度の旋律の抜粋を用いており，本実験の結果と直接比較することはできない．

(5) リコーダーの認識率が，個々の楽器レベルで 0.63%，カテゴリーレベルで 1.25%下り，また，オーボエのカテゴリーレベルの認識率が 0.67%下がった．しかし，リコーダーやオーボエはデータ数が少なく（それぞれ 160 個，151 個）多少の認識誤りがパーセンテージに大きく現れる．データ数で表せば，リコーダーの個々の楽器レベルで 1 個，カテゴリーレベルで 2 個，オーボエのカテゴリーレベルで 1 個，誤認識が増えたに過ぎない．

表 5 **5.** の実験結果（k-NN 法とベイズ決定規則との認識率の比較；個々の楽器レベルの認識率のみ）

楽器	k-NN 法 ($k=3$)			ベイズ決定規則		
記号	PCA1	PCA2	LDA	PCA1	PCA2	LDA
PF	53.94%	46.46%	63.39%	55.91%	59.06%	83.27%
CG	79.74%	77.16%	75.72%	98.28%	97.27%	90.23%
UK	94.58%	92.54%	97.63%	67.12%	80.00%	97.97%
AG	95.05%	92.79%	97.00%	19.97%	44.14%	83.93%
VN	47.73%	46.02%	45.83%	89.58%	84.47%	73.67%
VL	55.93%	54.24%	61.86%	71.19%	79.24%	76.27%
VC	86.20%	85.84%	84.23%	45.16%	30.82%	78.67%
TR	36.42%	38.41%	47.02%	41.72%	72.85%	82.12%
TB	70.99%	54.58%	77.86%	75.19%	78.24%	84.35%
SS	23.08%	14.20%	24.85%	48.52%	66.86%	65.89%
AS	37.59%	29.79%	40.43%	72.70%	41.84%	47.84%
TS	62.09%	66.01%	68.63%	30.07%	61.44%	66.01%
BS	68.84%	67.91%	66.98%	55.35%	54.42%	73.95%
OB	47.68%	48.34%	49.01%	43.71%	81.46%	72.19%
FG	64.10%	65.06%	74.36%	40.38%	30.12%	68.59%
CL	93.45%	87.93%	93.10%	95.51%	93.45%	92.07%
PC	84.08%	84.90%	84.08%	63.27%	58.37%	81.63%
FL	88.06%	72.39%	94.03%	35.82%	84.33%	85.07%
RC	97.50%	93.75%	97.50%	85.00%	96.25%	91.25%
平均	70.27%	66.98%	72.53%	62.11%	66.50%	79.73%

PCA1：主成分分析のみを用いて 79 次元に圧縮した場合．
PCA2：主成分分析のみを用いて 18 次元に圧縮した場合．
LDA：主成分分析を用いて 79 次元に圧縮し，さらに線形判別分析で 18 次元に圧縮した場合．

5. k-NN 法との比較

本章では，F0 依存多次元正規分布を仮定してベイズ決定規則を用いた場合（提案手法）と他の手法（ノン・パラメトリックな手法）を用いた場合，および，線形判別分析を用いた場合と用いなかった場合とで，認識率を比較する．なお，ノン・パラメトリックな手法としては，k-NN 法 ($k=3$) を取り上げた．

実験方法は **4.** と同じく，表 1 のデータ（総数：6247 個）を使ってクロスバリデーションを行う．実験結果（**表 5**）から以下の考察が得られる：

(1) 平均の認識率で，主成分分析・線形判別分析で次元圧縮した後に F0 依存多次元正規分布を仮定してベイズ決定規則を用いた場合（提案手法）が最も高かった（79.73%）．また，この方法は楽器毎の性能の偏りも最も小さかった．

(2) トランペット，ソプラノサックス，テナーサックス，オーボエ，フルートについて，主成分分析で 79

次元に圧縮してベイズ決定規則を用いた場合の認識率がいずれも低い（30.07%～48.52%）のに対して，主成分分析による次元圧縮を18次元にすると，認識率に大幅の改善が見られた（66.84%～84.33%）．これは，79次元正規分布のパラメータを推定するのに十分な数の学習データがなかったため（いずれも170個未満）と考えられる．しかし，全体では62.11%から66.50%に改善されたにすぎない．これは，主成分分析が識別を考慮した次元圧縮ではないため，識別に効果的な特徴が落とされる可能性があるからと考えられる．それに対し，線形判別分析はクラス内分散・クラス間分散最大化に基づく識別を考慮した次元圧縮法で，実際に認識率は79.73%と大幅に改善された．

(3) 本稿では，線形判別分析のみを用いて次元圧縮した場合については実験しなかった．これは，線形判別分析で用いる逆行列は，特徴空間に相関性の高い軸が含まれていると誤差が大きくなるため，線形判別分析による部分空間が，正常に算出されないためである．主成分分析は，特徴空間の次元を圧縮するだけでなく，各軸が直交するように空間を変形する．そのため，線形判別分析を適用する前に，主成分分析を用いて各軸を直交化することが有効である．

6. おわりに

本稿では，音高による音色変化の表現方法として，F0依存多次元正規分布を提案し，この分布を仮定した場合の識別関数をベイズ決定規則から定式化した．これは，各クラスのパターンが，音高によって平均が変化する多次元正規分布に従うと考え，音高による音色変化を基本周波数の関数，音高以外の要因による音色変化を共分散行列によって表すものである．本手法を実装・実験した結果，個々の楽器レベルで平均16.48%，カテゴリーレベルで平均20.67%の認識誤りを削減することができた．

本稿で提案したF0依存多次元正規分布は，ベイズ決定規則への応用のみに限定されるものではない．今後は，この枠組みを応用して，より高性能な識別手法の設計に取り組むとともに，より多くの楽器に対応できるよう，他の特徴量の導入も検討する．さらに，混合音への適用などにも取り組んでいく予定である．

謝辞 本研究は，日本学術振興会から交付された科学研究費補助金およびNTTコミュニケーション科学基礎研究所から援助を受けた．また，本研究の実験において，文献15)の「RWC研究用音楽データベース：楽器音」(RWC-MDB-I-2001)を使用した．

参考文献

1) P. Cosi, G. D. Poli and G. Lauzzana: "Auditory Modelling and Self-Organizing Neural Networks for Timbre Classification", *J. New Music Research*, **23**, pp.71–98, 1994.
2) G. J. Brown and M. Cooke: "Perceptual Grouping of Musical Sounds: A Computational Model", *J. New Music Research*, **23**, pp.107–132, 1994.
3) 柏野 邦夫，中臺 一博，木下 智義，田中 英彦："音楽情景分析の処理モデルOPTIMAにおける単音の認識"，信学論，**J79-D-II**, 11, pp.1751–1761, 1996.
4) 柏野 邦夫，村瀬 洋："適応型混合テンプレートを用いた音源同定"，信学論，**J81-D-II**, 7, pp.1510–1517, 1998.
5) 木下 智義，坂井 修一，田中 英彦："周波数成分の重なり適応処理を用いた複数楽器の音源同定処理"，信学論，**J83-D-II**, 4, pp.1073–1081, 2000.
6) K. D. Martin: "Sound-Source Recognition: A Theory and Computional Model", PhD Thesis, MIT, 1999.
7) I. Fujinaga and K. MacMillan: "Realtime Recognition of Orchestral Instruments", *Proc. of ICMC*, 2000.
8) A. Eronen: "Automatic Musical Instrument Recognition", M.Sc. Thesis, Tampere Univ. of Tech., 2001.
9) 山口 公典，安藤 繁雄："短時間スペクトル分析法の自然楽器音への適用"，音響誌，**33**, 6, pp.291–300, 1977.
10) 早坂 寿雄："楽器の科学"，電子情報通信学会，1992.
11) H. F. Olson（平岡 正徳訳）："音楽工学"，誠文堂新光社，1969.
12) 安藤 由典："楽器の音響学"，音楽之友社，1996.
13) A. S. Bregman: "*Auditory Scene Analysis*", MIT Press, 1990.
14) 北原 鉄朗，後藤 真孝，奥乃 博："音高による音色変化に着目した音源同定手法"，情処研報，2001-MUS-40, pp.7–14, 2001.
15) 後藤 真孝，橋口 博樹，西村 拓一，岡 隆一："RWC研究用音楽データベース：音楽ジャンルデータベースと楽器音データベース"，情処研報，2002-MUS-45, pp.19–26, 2002.
16) 原 祐一郎，井口 征士："複素スペクトルを用いた周波数同定"，計測論，**19**, 9, pp.718–723, 1983.
17) 後藤 真孝，村岡 洋一："打楽器音を対象にした音源分離システム"，信学論，**J77-D-II**, 5, pp.901–911, 1994.
18) A. Savitzky and M. J. E. Golay: "Smoothing and Differentiation of Data by Simplified Least Squares Procudures", *Anal. Chem.*, **36**, 8, pp.1627–1639, 1964.
19) C. Roads（青柳 龍也 他訳）："コンピュータ音楽――歴史・テクノロジー・アート――"，東京電機大学出版局，2001.
20) 片寄 晴弘："自動採譜"，信学誌，**79**, 3, pp.287–289, 1996.
21) R. A. Rasch and R. Plomp（宮坂 栄一訳）："楽音の知覚"，音楽の心理学（上）第1章，西村書店，1987.
22) J. C. Risset and D. L. Wessel（宮坂 栄一訳）："分析と合成による音色の探求"，音楽の心理学（上）第2章，西村書店，1987.
23) K. W. Berger: "Some Factors in the Recognition of Timbre", *J. Acoust. Soc. Am.*, **36**, 10, pp.1888–1891, 1964.