

# 音高による音色変化に着目した楽器音の音源同定： F0 依存多次元正規分布に基づく識別手法

北原 鉄 朗<sup>†</sup> 後藤 真 孝<sup>††</sup> 奥 乃 博<sup>†</sup>

本論文では、音高による音色変化を考慮する楽器音の音源同定手法を提案する。楽器音の音色が音高によって変化することは、従来から広く知られているにもかかわらず、これを適切に扱える音源同定手法については、研究されてこなかった。本論文では、音高による音色変化を適切に扱うため、平均が基本周波数によって変化する多次元正規分布を提案する。そして、音色空間（楽器音の特徴空間）上で各楽器音データがこの分布に従うと仮定し、この分布のための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、音高による音色変化を考慮しない多次元正規分布を用いた場合の誤認識全体のうち、個々の楽器レベルでは 16.48%、カテゴリーレベルでは 20.67%の誤認識を削減することができた。

## Musical Instrument Identification Considering Pitch-dependent Characteristics of Timbre: A Classifier Based on F0-dependent Multivariate Normal Distribution

TETSURO KITAHARA,<sup>†</sup> MASATAKA GOTO<sup>††</sup> and HIROSHI G. OKUNO<sup>†</sup>

This paper describes a musical instrument identification method that takes pitch-dependent characteristics of timbre into consideration. Although it is known that timbres of musical instruments have pitch-dependent characteristics, the characteristics were not dealt with in most of the previously proposed musical instrument identification methods. To deal with the characteristics, we propose *F0-dependent multivariate normal distribution* and 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

楽器音の音源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ものである<sup>1)~8)</sup>。そのため、音声認識や文字認識に比べ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<sup>9)~12)</sup>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<sup>3)</sup>（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<sup>13)</sup>）の重要な要素技術の1つと位置づけ、(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。このように2段階のアプローチをとるのは、音源同定は単音であっても難しい問題だからである。たとえば、文献6)では、音楽経験者を対象とした聴取実験（14楽器の単音を聞いて、その楽器名を27個の楽器名が書かれたリストから選ぶという実験）で、45.9%の正解率が報告されている。

楽器音の音源同定の精度を高くするのが難しい原因は、楽器の音色が、楽器の個体差や音高などのさまざまな要因により変化するからである。しかし、従来の音源同定研究の多く<sup>1)~3),5)~8)</sup>は、こうした音色変化の問題を明示的に扱っていなかった。それに対し、柏野らの適応型混合テンプレート法<sup>4)</sup>は、テンプレートフィルタリングにより楽器の個体差を、位相トラッキングにより音高の揺らぎを吸収することで、音色変化

<sup>†</sup> 京都大学大学院情報学研究科知能情報学専攻  
Department of Intelligence Science and Technology,  
Graduate School of Informatics, Kyoto University

<sup>††</sup> 科学技術振興事業団さきかけ研究 21「情報と知」領域/産業技術総合研究所

“Information and Human Activity”, PRESTO, JST/  
National Institute of Advanced Industrial Science and  
Technology

の問題に対処していた。ただし、テンプレートを各楽器で半音ごとに用意しているものの、音高が変化すると音色がどのように変化するかを効率的にモデル化する手法は扱われていなかった。一般に楽器の音域は広く、たとえば、ピアノの音域は7オクターブにもわたる。そのため、高い音域と低い音域では音色が大きく異なり、音高を考慮しなければ、各楽器の全音域で適切に音源同定するのが困難となる。しかも楽器の音色は、通常、低域から高域にいくに従って連続的に変わることを、音源同定手法は考慮しなければならない。

本論文では、音高による音色変化を効率的にモデル化する手法として F0 依存多次元正規分布と、それに基づく音源同定手法を提案する。これは、音高が物理量（基本周波数）とほぼ 1 対 1 で対応することに着目して、音高による音色変化を基本周波数の関数として表現するものである。具体的には、音色空間（楽器音の特徴空間）上で、各楽器音データは、基本周波数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）に従うと仮定し、この分布のパラメータ推定法を提案する。そして、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する。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<sup>1)~8)</sup>では議論されてこなかった。

以下、まず 2 章で 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 章で提案手法の処理の流れを述べ、4 章で評価実験について述べる。さらに、5 章でベイズ決定規則と  $k$ -NN 法を比較し、最後に 6 章でまとめをする。

## 2. F0 依存多次元正規分布

本論文の音源同定方式では、各楽器名がラベルづけられた楽器音の特徴ベクトルのデータベース（個々の特徴ベクトル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<sup>10)</sup>。
- (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<sup>10)</sup>。
- (3) 一部の楽器では音高により発音体が異なり、各

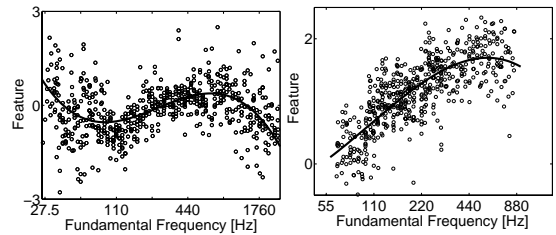


図1 代表値関数（太字）の例。左の図は線形近似では精度が不十分な例（ピアノの第4軸）で、右の図は音高による音色変化が特に顕著な例（チェロの第1軸）である

Fig.1 Examples of representation functions.

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、入力（同定対象）と同じ音高の学習パターンのみを使って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。

本論文では、学習パターンの分布の音高依存性が主に平均に顕著に現れる（付録参照）ことに着目し、平均を基本周波数の関数として拡張した多次元正規分布を提案する。これにより、音高による音色変化を考慮しながら、全音域の学習パターンを1つの分布で表現することができる。本論文では、このように拡張された多次元正規分布を 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（平均と共分散）の推定法を述べる。

### 2.1 代表値関数

基本周波数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）。この近似曲線を代表値関数と呼び、 $\mu_i(f)$  と書く（ $i$ : 楽器名）。ここでは、近似関数として3次関数を用いる。これは、複雑な音色変化を表現できることと、少ないデータからでも推定できることを両立できる次数を、予備実験により求めたものである。

### 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法を述べる。F0 依存多次元正規分布の共分散行列は、代表値関数からのちらばりの程度を表す。代表値関数は音高による音色変化を表すので、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られる。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。本論文では、この F0 依存多次元正規分布における共分散行列を F0 正規化共分散行列と

呼び、 $\Sigma_i$  と書く。

### 2.3 ベイズ決定規則による識別

各楽器  $\omega_i$  の学習パターンが、F0 依存多次元正規分布に従うと仮定し、事後確率  $p(\omega_i|\mathbf{x}; f)$  が最大になる楽器名を同定結果として出力する。ここで、パラメータ  $f$  は入力パターン  $\mathbf{x}$  の基本周波数で、本研究で新たに導入されたものである。

ベイズの定理より、事後確率は、確率密度関数  $p(\mathbf{x}|\omega_i; f)$  と事前確率  $p(\omega_i; f)$  を使って、

$$p(\omega_i|\mathbf{x}; f) = \frac{p(\mathbf{x}|\omega_i; f)}{p(\mathbf{x}; f)} p(\omega_i; f)$$

と表されるので、識別規則は次のように定式化される：

$$\begin{aligned} \hat{\omega} &= \operatorname{argmax}_{\omega_i} p(\omega_i|\mathbf{x}; f) \\ &= \operatorname{argmax}_{\omega_i} \frac{p(\mathbf{x}|\omega_i; f)}{p(\mathbf{x}; f)} p(\omega_i; f) \\ &= \operatorname{argmax}_{\omega_i} p(\mathbf{x}|\omega_i; f) p(\omega_i; f) \\ &= \operatorname{argmax}_{\omega_i} \{ \log p(\mathbf{x}|\omega_i; f) + \log p(\omega_i; f) \}. \end{aligned} \quad (1)$$

また、F0 依存多次元正規分布の確率密度関数は

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

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あり、その定義は

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

で与えられる（' は転置）。この式を式 (1) に代入することで、次の識別規則が得られる：

$$\hat{\omega} = \operatorname{argmax}_{\omega_i} \left\{ -\frac{1}{2} D^2(\mathbf{x}, \boldsymbol{\mu}_i(f)) - \frac{1}{2} \log |\Sigma_i| - \frac{d}{2} \log 2\pi + \log p(\omega_i; f) \right\}.$$

ここで、事前確率について述べる。本論文では、事前確率を、音域的に発音可能かどうかに基づいて定める。楽器  $\omega_i$  の音域を  $R_i$  とすると、楽器  $\omega_i$  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、識別対象の楽器のうち、基本周波数  $f$  で発音可能な楽器の個数を表す。これにより、発音不可能な音域では事前確率が 0 になるため、識別結果は、音域的に発音可能な楽器に限定される。

## 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ものを 129 個抽出する。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楽器のパターンが 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：4,096 点、時間分解能：10 ms）、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<sup>14)</sup>をハニング窓用に拡張したもの<sup>15)</sup>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、音高（C4 など）を人手で与え、その音高に対応する周波数（平均律で算出）の近傍（200 cent 以内）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<sup>3),6)</sup>や楽器音響学・楽器物理学などの知見<sup>9)~12)</sup>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（ $i = 2, 3, \dots, 29$ ）、
- 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 音が鳴り続けている時間（周波数成分全体のパワーがしきい値を超えている時間）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超えている時間）が  $p\%$  である高調

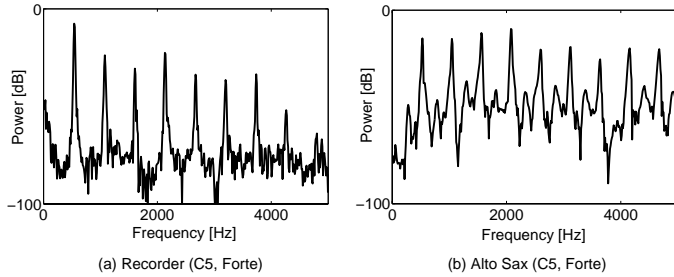


図3 発音開始直後のパワースペクトル。(b)のように非高調波成分が多く含まれていると、(a)に比べて高調波成分のピークの尖度が低くなるのが分かる

Fig. 3 Power spectrum at onset time.

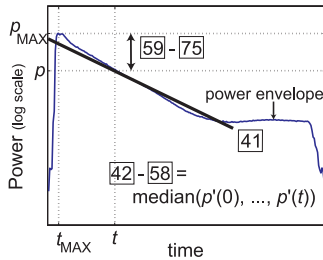


図2 パワーの時間変化に関する特徴(ピアノ, C4, 強).  $p'(t)$  はパワー包絡線の時刻  $t$  の微分係数を表す

Fig. 2 Temporal features of power.

波成分の個数 ( $p = 10, 20, \dots, 90$ ).

(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(図2)。

[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 ( $t = 0.15, 0.20, \dots, 0.95$ ),

[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 ( $t = 0.15, 0.20, \dots, 0.95$ ) .

(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<sup>16)</sup>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 ( $k = 1, 2, \dots, 13$ ),

[95]~[107]  $k$  次のMFCCの時間変化の振動数 ( $k = 1, 2, \dots, 13$ ) .

(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  $f[\text{Hz}]$  とすると、 $0.75f[\text{Hz}]$  から  $1.5f[\text{Hz}]$  までの範囲)の部分を取り出す。そして、切り出された各ピーク付近の尖度を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(図3)。そこで、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

上記の特徴量を平均が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、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。本論文では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

これらの処理は、すべての学習データを用いて行われる。これは、学習データが少ないと、基本周波数ごとに信頼性のある特徴空間を得るのが困難だからである。

表 1 使用した楽器音データベースの内訳

Table 1 Contents of musical instrument sound database used in this paper.

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			666
15	バイオリン (VN)	3	G3-E7	そ れ ぞ れ	通 常 の 中 奏 法 の み	528
16	ビオラ (VL)	3	C3-F6	そ れ ぞ れ	通 常 の 中 奏 法 の み	472
17	チェロ (VC)	3	C2-F5	そ れ ぞ れ	通 常 の 中 奏 法 の み	558
21	トランペット (TR)	2	E3-A#6	強	3	151
22	トロンボーン (TB)	3	A#1-F#5	中	種 類	262
25	ソプラノサックス (SS)	3	G#3-E6	弱		169
26	アルトサックス (AS)	3	C#3-A5	弱		282
27	テナーサックス (TS)	3	G#2-E5	弱		153
28	バリトンサックス (BS)	3	C2-A4	弱		215
29	オーボエ (OB)	2	A#3-G6			151
30	ファゴット (FG)	3	A#1-D#5			312
31	クラリネット (CL)	3	D3-F6			263
32	ピッコロ (PC)	3	D5-C8			245
33	フルート (FL)	2	C4-C7			134
34	リコーダー (RC)	3	C4-B6			160

\* 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

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。

### 3.4 識 別

2章で述べたように、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、各楽器のパターンが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<sup>17)</sup>を使用する。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, 収録時のノイズが大きいものなどを除いた19種類の楽器を使用する。使用したデータ(総数:6,247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で評価するという実

表 2 19 楽器の分類

Table 2 Categorization of 19 instruments.

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, リコーダー

験を繰り返して、認識率の平均を求める。そのため、代表値関数やF0正規化共分散行列の算出では、表1のデータ数の90%のデータを用いることになる。

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、カテゴリーレベルで両者を区別することはできる。

本論文では、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。この分類は、楽器の発音機構や従来研究<sup>6),8)</sup>に基づいて本研究で定義したものである(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<sup>8)</sup>)。まず、同定対象を個々の楽器レベルで(すなわち楽器名を)同定する。そして、この同定結果と正解とがカテゴリーレベルで一致しているかどうかを表2の分類に基づいて決定し、カテゴリーレベルの認識率を算出する。

### 4.2 実験結果

通常多次元正規分布を仮定して識別した場合(音高による音色変化を考慮しない場合)と、F0依存多次元正規分布を仮定して識別した場合(提案手法)の両方の認識率を表3に示す。本論文で提案する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%改善された。これは、音高による音色変化を考慮しない場合の誤認識全体を1とすると、その16.48%を削減できた

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)

Table 3 Accuracy by F0-independent distribution (baseline) and F0-dependent distribution (proposed).

楽器記号	個々の楽器レベル			カテゴリーレベル		
	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)

ことを意味する。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% 改善された。これは、音高による音色変化を考慮しない場合の認識率全体を 1 とすると、その 20.67% を削減できたことを意味する。

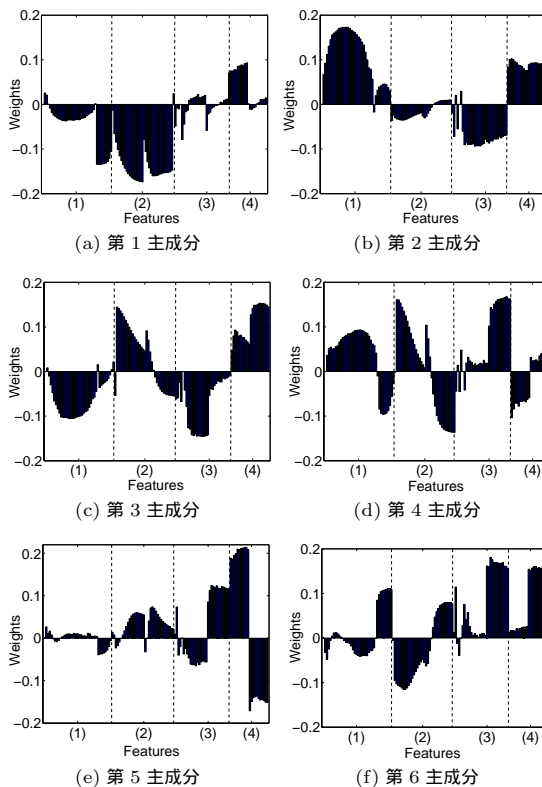
これらの結果が有意であることを  $t$  検定 (片側検定) で示す。各楽器における両手法の認識率の差を  $d_i$  ( $i = 1, \dots, n$ ) とすると、検定統計量  $t_0$  は、

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

で与えられる。ここで、 $\bar{d}$  は  $d_1, \dots, d_n$  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 4 に示す。図 4 (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いわれており、このことを裏づける結果になったといえる。そ



注 図の横軸は、3.2 節の 129 個の特徴量に対応し、図中の (1) ~ (4) は、3.2 節の特徴量の説明における (1) ~ (4) に対応する

図 4 主成分分析による各主成分の重み値

Fig. 4 Excerpts of weights by principal component analysis.

のほか、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<sup>12)</sup>、ほとんど考慮さ

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<sup>18)</sup>。たとえば、安藤は、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」と述べている<sup>12)</sup>。

表 4 特徴空間変形における基底ベクトルの一部

Table 4 Excerpts of basic vectors in transformation of feature space.

	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)	

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 と高く現れたほかは、32~40 (音が鳴り続け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<sup>19),20)</sup>、上記の結果は、これを裏づ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 150 ms 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<sup>12),19),20)</sup>、楽器音の音響信号を逆転再生すると音源同定能力が低下するという実験結果<sup>21)</sup>などもこのことを裏づけている。

##### (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 0.2607~0.5400)。これは、4.3 節でも述べたように、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、
- 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、
- 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、ということが知られている<sup>10),12)</sup>。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二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見られる。

(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% を超えており、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、サクスの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、ソプラノサクスの7.1%、アルトサクスの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%下がり、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)、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、誤認識が増えたにすぎない。

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章と同じく、表1のデータ(総数:6,247個)を使ってクロスバリデーションを行う。実験結果(表5)から以下の知見が得られる:

(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(f)が最も高かった(79.73%)。また、この方法は楽器ごとの性能の偏りも最も小さかった。

(2) 表5の(d)と(e)を比較すると、(e)(18次元特徴空間;累積寄与率88%)の認識率が(d)(79次元特徴空間;累積寄与率99%)よりも4.39%高かった。ほかにもさまざまな次元数で実験したところ、23次元(累積寄与率:90%)のときが最も認識率が高く(69.09%)、それより次元数を高くすると認識率は低くなる現象がみられた(図5)。このように、ある次元数を超えると次元数を上げることで認識率が下がる現象(ヒューズの現象<sup>22)</sup>)が、本研究でも観測された。

表5 5章の実験結果(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)

Table 5 Accuracy by k-NN rule and Bayes decision rule.

	(a)	(b)	(c)	(d)	(e)	(f)
	k-NN法(k=3)			ベイズ決定規則		
	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。

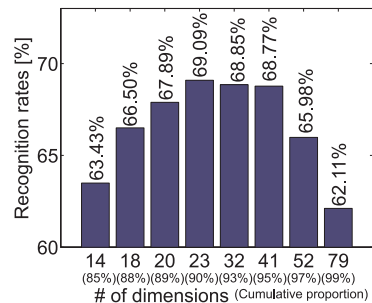


図5 主成分分析のみを用いて次元圧縮したときの次元数と認識率の関係。23次元以下では、次元数を上げると認識率も向上するが、23次元を超えると、次元数を上げることで認識率は逆に下がることが分かる。ただし、いずれの場合も線形判別分析を併用した場合の認識率(79.73%)より低い

Fig. 5 Relationship between the number of dimensions and recognition rates.

(3) 楽器別にみると、トランペット、ソプラノサクスの、テナーサクスの、オーボエ、フルートの5つの楽器において、(d)の認識率がいずれも低い(30.07~48.52%)のに対して、(e)では認識率が大幅に改善された(61.44~84.33%)。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



かったためと考えられる(いずれも170個未満). 文献23)では, 学習データは特徴空間の次元数の5~10倍程度あるのが望ましいと述べられており, 本実験結果はこれと一致する.

(4) 表5のすべてにおいて, 線形判別分析を併用した方(f)が認識率は高かった(79.73%). これは, 線形判別分析が, クラス内分散・クラス間分散比最大化基準に基づく, 識別を考慮した次元圧縮法だからである.

(5) 本論文では, 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

本論文では, 音高による音色変化を考慮する音源同定手法として, F0依存多次元正規分布に基づく識別手法を提案した. この手法は, 各楽器音データが, 基本周波数によって平均が変化する多次元正規分布に従うと仮定し, 基本周波数によって変化する平均を関数近似により求めるものである. 実験の結果, 音高による音色変化を考慮しない場合の誤認識全体のうち, 個々の楽器レベルで16.48%, カテゴリーレベルで20.67%の誤認識を削減することができた.

本論文で提案した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 本研究の一部は, 日本学術振興会科学研究費補助金基盤研究(B)第12480090号, NTTコミュニケーション科学基礎研究所, および21世紀COEプログラム「知識社会基盤構築のための情報学拠点形成」によるものである. また, 本研究の実験において, 文献17)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. 最後に, ご討論いただいた麻生英樹氏(産業技術総合研究所), 中臺一博氏(科学技術振興事業団; 現在, 株式会社ホンダ・リサーチ・インスティテュート・ジャパン)に感謝する.

## 参考文献

- 1) Cosi, P., Poli, G.D. and Lauzzana, G.: Auditory Modelling and Self-Organizing Neural Networks for Timbre Classification, *J. New Music Research*, Vol.23, pp.71-98 (1994).
- 2) Brown, G.J. and Cooke, M.: Perceptual Grouping of Musical Sounds: A Computational Model, *J. New Music Research*, Vol.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, Vol.J79-D-II, No.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, Vol.J81-D-II, No.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, Vol.J83-D-II, No.4, pp.1073-1081 (2000).
- 6) Martin, K.D.: Sound-Source Recognition: A Theory and Computational Model, Ph.D. Thesis, MIT (1999).
- 7) Fujinaga, I. and MacMillan, K.: Realtime Recognition of Orchestral Instruments, *Proc. ICMC*, pp.141-143 (2000).
- 8) Eronen, A.: 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, Vol.33, No.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) Olson, H.F. (著), 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) Bregman, A.S.: *Auditory Scene Analysis*, MIT Press (1990).
- 14) 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, Vol.19, No.9, pp.718-723 (1983).
- 15) 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, Vol.J77-D-II, No.5, pp.901-911 (1994).
- 16) Savitzky, A. and Golay, M.J.E.: Smoothing and Differentiation of Data by Simplified Least Squares Procedures, *Anal. Chem.*, Vol.36, No.8, pp.1627-1639 (1964).
- 17) 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, 情報処理学会研究報告, 2002-MUS-45, pp.19-26 (2002).

- 18) Roads, C. (著), 青柳龍也ほか(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).
- 19) Rasch, R.A. and Plomp, R. (著), 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).
- 20) Risset, J.C. and Wessel, D.L. (著), 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).
- 21) Berger, K.W.: Some Factors in the Recognition of Timbre, *J. Acoust. Soc. Am.*, Vol.36, No.10, pp.1888–1891 (1964).
- 22) Hughes, G.F.: On the Mean Accuracy of Statistical Pattern Recognition, *IEEE Trans. Inf. Theory*, Vol.14, pp.55–63 (1968).
- 23) Jain, A.K. and Chandrasekaran, B.: Dimensionality and Sample Size Considerations in Pattern Recognition, *Handbook of Statistics 2: Classification, Pattern Recognition and Reduction of Dimensionality*, Krishnaiah, P.R. and Kahal, L.N. (Eds.), North-Holland, pp.835–855 (1982).

## 付 録

### A.1 共分散行列の F0 依存性について

ここでは, 共分散行列が平均に比べて, 基本周波数によってあまり変化しないことを確認する. ただし, ここでは共分散行列の対角成分(実際には標準偏差)のみを扱う. これは, 主成分分析によって非対角成分が対角成分に比べて十分に小さくなるためである.

平均と標準偏差の F0 依存性は, 次の手順で調べる:

- (1) 各楽器に対して, 対象音域の低い方から 1 オクターブの区間(ピアノなら A0–G $\sharp$ 1)に含まれるデータを使って, 各特徴量の平均と標準偏差を求める.
- (2) 半音ずつずらして(ピアノなら A $\sharp$ 0–A1, B0–A $\sharp$ 1, ...) 同様の計算を行う.
- (3) このようにして得られた各楽器・各特徴量の平均と標準偏差が, 基本周波数によってどの程度変化するかをとらえるため, それぞれの基本周波数方向の分散を計算する.
- (4) この値が大きければ, 基本周波数による変動が大きいことを示す.

この手順により得られた結果を表 6 に示す. 表より, 多くの楽器・特徴量において, 標準偏差の基本周波数による変動が, 平均のそれより小さいことが分かる. 表では紙面の都合により第 7 軸～第 18 軸については省略したが, 第 1 軸～第 6 軸とほぼ同様の結果が得られている.

表 6 各楽器・各特徴量における平均と標準偏差の基本周波数による変動の比較(抜粋). 標準偏差の基本周波数による変動(下段)が, 平均のそれ(上段)よりもかなり小さいことが分かる

Table 6 Comparison of the F0 dependency of mean and standard deviation.

	第 1 軸	第 2 軸	第 3 軸	第 4 軸	第 5 軸	第 6 軸
PF	0.5425	0.1438	0.3782	0.1195	0.1408	0.1243
	0.0343	0.0192	0.0075	0.0352	0.0077	0.0095
CG	1.1160	0.0631	0.0569	0.1168	0.0159	0.0669
	0.0163	0.0050	0.0014	0.0037	0.0077	0.0018
UK	0.1149	0.0129	0.0737	0.0171	0.0365	0.0024
	0.0133	0.0072	0.0318	0.0142	0.0183	0.0020
AG	0.1365	0.1412	0.0430	0.0948	0.0613	0.0326
	0.0309	0.0021	0.0010	0.0047	0.0025	0.0020
VN	0.0962	0.3119	0.0171	0.0190	0.2122	0.0813
	0.0055	0.0619	0.0017	0.0134	0.0413	0.0209
VL	0.0226	0.0087	0.0043	0.0024	0.0829	0.0113
	0.0005	0.0073	0.0014	0.0021	0.0017	0.0015
VC	0.1179	0.0522	0.0367	0.0694	0.4037	0.0253
	0.0026	0.0304	0.0031	0.0033	0.0201	0.0037
TR	0.0214	0.0337	0.0009	0.0093	0.0621	0.0104
	0.0044	0.0025	0.0008	0.0034	0.0057	0.0019
TB	0.0305	0.1097	0.0142	0.4370	0.6630	0.0547
	0.0245	0.0173	0.0075	0.0272	0.0173	0.0069
SS	0.0108	0.0090	0.0033	0.0179	0.0620	0.0126
	0.0020	0.0104	0.0010	0.0006	0.0028	0.0003
AS	0.0057	0.0185	0.0322	0.1568	0.0103	0.0049
	0.0011	0.0224	0.0055	0.0099	0.0052	0.0026
TS	0.0051	0.0338	0.1265	0.1024	0.0316	0.0048
	0.0026	0.0037	0.0213	0.0058	0.0032	0.0084
BS	0.0119	0.0656	0.0247	0.5389	0.1487	0.0414
	0.0096	0.0334	0.0186	0.0638	0.0178	0.0100
OB	0.0027	0.0158	0.0130	0.0037	0.0008	0.0069
	0.0003	0.0038	0.0009	0.0005	0.0046	0.0008
FG	0.0344	0.0317	0.0364	0.0579	0.0205	0.0071
	0.0075	0.0091	0.0038	0.0178	0.0038	0.0088
CL	0.0014	0.1213	0.0089	0.0297	0.0131	0.0060
	0.0011	0.0030	0.0014	0.0027	0.0181	0.0010
PC	0.0752	0.0013	0.1545	0.1782	0.0629	0.0364
	0.0117	0.0338	0.0064	0.0017	0.0105	0.0152
FL	0.0100	0.0119	0.0066	0.0097	0.0090	0.0374
	0.0005	0.0016	0.0011	0.0025	0.0011	0.0011
PC	0.0024	0.0214	0.0187	0.0032	0.0099	0.1358
	0.0005	0.0006	0.0055	0.0017	0.0015	0.0103

(平成 14 年 9 月 17 日受付)

(平成 15 年 9 月 5 日採録)



北原 鉄朗(学生会員)

2002 年東京理科大学理工学部情報科学科卒業。現在, 京都大学大学院情報学研究科知能情報学専攻修士課程在学中。音楽情報処理に興味を持つ。電子情報通信学会, 人工知能学会, 日本音響学会, 日本音楽知覚認知学会学生会員。



後藤 真孝 (正会員)

1993年早稲田大学理工学部電子通信学科卒業。1998年同大学大学院理工学研究科博士後期課程修了。同年、電子技術総合研究所(2001年に独立行政法人産業技術総合研究所に改組)

に入所し、現在に至る。2000年から2003年まで科学技術振興事業団さきがけ研究21研究員を兼任。博士(工学)。音楽情報処理、音声言語情報処理等に興味を持つ。1992年JUS設立10周年記念UNIX国際シンポジウム論文賞, 1993年NICOGRAPH'93 CG教育シンポジウム最優秀賞, 1997年情報処理学会山下記念研究賞(音楽情報科学研究会), 1999年電気関係学会関西支部連合大会奨励賞, 2000年WISS2000論文賞・発表賞, 2001年日本音響学会第18回粟屋潔学術奨励賞・第5回ポスター賞, 2002年情報処理学会山下記念研究賞(音声言語情報処理研究会), 2002年日本音楽知覚認知学会研究選奨, 2003年インタラクシオン2003ベストペーパー賞各受賞。電子情報通信学会, 日本音響学会, 日本ソフトウェア科学会, 日本音楽知覚認知学会, ISCA各会員。



奥乃 博 (正会員)

1972年東京大学教養学部基礎科学科卒業。日本電信電話公社, NTT, 科学技術振興事業団北野共生システムプロジェクト, 東京理科大学理工学部情報科学科を経て, 2001年4月

より京都大学情報学研究科知能情報学専攻教授。博士(工学)。この間, スタンフォード大学客員研究員, 東京大学工学部客員助教授。人工知能, 音環境理解, ロボット聴覚の研究に従事。1990年度人工知能学会論文賞, IEA/AIE-2001最優秀論文賞, IEEE/RSJ IROS-2001 Best Paper Nomination Finalist, 第2回船井情報科学振興賞受賞。本学会英文図書委員。著編書: 『インターネット活用術』(岩波書店, 1996), “Computational Auditory Scene Analysis”(共編, LEA, 1998), “Advanced Lisp Technology”(共編, Taylor & Francis, 2002)等。