音色空間の音高依存性を考慮した楽器音の音源同定

◎ 北原 鉄朗 (京大)，後藤 真孝 (科技団さきがけ研究 21 / 産総研)，奥乃 博 (京大)

1. はじめに

楽器音の音源同定では，音高による音色変化を適切に扱うことが重要である．楽器音は，音声に比べて広い音域を持つため，音高による音色変化が顕著に現れるからである．しかし，従来研究[1)~5)]では，音高による音色変化はほとんど扱われてこなかった．

本稿では，音高による音色変化を考慮するため，**F0** 依存多次元正規分布を提案する．これは，音色空間 (楽器音の特徴空間) 上における楽器音データの分布が音高によって変化すると考え，この音高依存性を基本周波数の関数として表すものである．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．

2. F0 依存多次元正規分布

本稿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求める．そして，入力に対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．

本稿では，各楽器の学習パターンが，平均が基本周波数によって連続的に変化し，共分散行列が基本周波数に依存しない多次元正規分布に従うと仮定する．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 (平均と共分散行列) 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，$\boldsymbol{\mu}_i(f)$ と書く (i :楽器番号)．ここでは，近似関数として 3 次関数を用いる．

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出法を述べる．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．本稿では，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，Σ_i と書く．

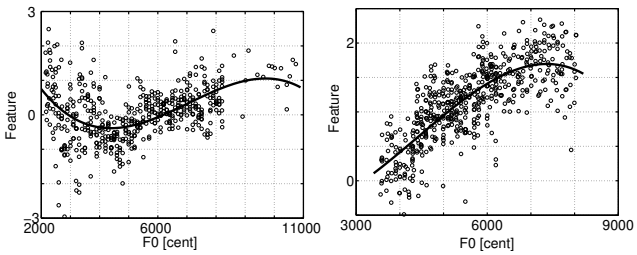

図 1 代表値関数 (太字) の例．左の図は線形近似では精度が不十分な例 (ピアノの第 4 軸) で，右の図は音高による音色変化が特に顕著な例 (チェロの第 1 軸) である．

Musical Instrument Identification Considering Pitch-dependent Characteristics of Timbre Space
by Tetsuro Kitahara (Kyoto Univ.), Masataka Goto (PRESTO, JST / AIST) and Hiroshi G. Okuno (Kyoto Univ.)

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．各楽器ω_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，この分布の確率密度関数 $p(\boldsymbol{x}|\omega_i; f)$ と事前確率 $p(\omega_i; f)$ とを使って，パターン$\boldsymbol{x}$が入力されたときの識別関数を次式で定義する：

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

ここで，パラメータ f は入力パターン$\boldsymbol{x}$の基本周波数で，本研究で新たに導入されたものである．また，F0 依存多次元正規分布の確率密度関数は次式で与えられる：

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

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)．そこで，この式を式 (1) に代入することで次の識別関数 g が得られる：

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

この $g_i(\boldsymbol{x}; f)$ を最大にする楽器名 ($k = \mathrm{argmax}_i\, g_i(\boldsymbol{x}; f)$ とおいたときのω_k) を同定結果として出力する．

本稿では，事前確率 $p(\omega_i; f)$ を，入力パターンの基本周波数が f であるときに，各楽器の音域に f が含まれているかどうかに基づいて定める．すなわち，楽器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

提案手法の処理の流れを以下に示す：

(1) 調波構造推定

まず，各時刻 (10ms ごと) において短時間フーリエ変換によりパワースペクトルを計算し，その周波数方向のピークを抽出する．そして，それらのピークから調波構造 (30 次まで) を求める．ただし，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 (200cent 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．

表 1　129 個の特徴量の概要

(1)	スペクトルに関する定常的特徴 (40 個) 周波数重心の時間方向の中央値，他．
(2)	パワーの時間変化に関する特徴 (35 個) パワー包絡線の線形最小二乗法による近似直線の傾き，他．
(3)	各種変調の振幅 / 振動数 (32 個) 振幅変調，周波数変調，周波数重心の時間変化，MFCC の時間変化などの振幅 / 振動数．
(4)	発音開始直後のピーク尖度に関する特徴 (22 個) 発音開始直後 150ms 間における各高調波成分のピークの尖度を時間方向につないだものに対する，時間方向の平均値と時間変化の振幅．

表 2　使用したデータベースの内訳

楽器名 (楽器番号，楽器記号)	ピアノ (01, PF)，クラシックギター (09, CG)，ウクレレ (10, UK)，アコースティックギター (11, AG)，バイオリン (15, VN)，ビオラ (16, VL)，チェロ (17, VC)，トランペット (21, TR)，トロンボーン (22, TB)，ソプラノサックス (25, SS)，アルトサックス (26, AS)，テナーサックス (27, TS)，バリトンサックス (28, BS)，オーボエ (29, OB)，ファゴット (30, FG)，クラリネット (31, CL)，ピッコロ (32, PC)，フルート (33, FL)，リコーダー (34, RC)
楽器個体	TR, OB 以外 :3 種類．TR, OB :2 種類．
音の強さ	各楽器，強・中・弱の 3 種類．
奏法	通常の奏法のみ．
データ数*	PF: 508, CG: 696, UK: 295, AG: 666, VN: 528, VC: 558, TR: 151, TB: 262, SS: 169, AS: 282, TS: 153, BS: 215, OB: 151, FG: 312, CL: 263, PC: 245, FL: 134, RC: 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．

(2)　特徴抽出

先行研究や楽器音響学・楽器物理学などの知見を参考に定めた 129 個の特徴量 (表 1) を抽出する．

(3)　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 129 次元から 79 次元に圧縮される．その後，線形判別分析によりさらに次元を圧縮する．本稿では 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．

(4)　識　別

次元圧縮された 18 次元の特徴空間上で，各楽器のパターンが前述の F0 依存多次元正規分布に従うと仮定し，ベイズ決定規則から楽器名を同定する．

4. 評価実験

実楽器の単音データベース (DB) として，RWC 研究用音楽 DB の楽器音 DB (RWC-MDB-I-2001)[6] を使用する．この DB 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．使用するデータ (総数 :6247 個) の内訳を表 2 に示す．表 2 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．

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．

表 3　実験結果

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)
「カテゴリーレベル」の認識率とは，19 楽器を {PF}, {CG, UK, AG}, {VN, VL, VC}, {TR, TB}, {SS, AS, TS, BS}, {OB, FG}, {CL}, {PC, FL, RC} の 8 カテゴリーにまとめ，8 カテゴリーの同定問題とした場合の認識率である．

楽器別にみると，19 楽器のうち 6 つの楽器 (PF, TR, TB, SS, BS, FG) で 7%以上認識率が向上した．特に，ピアノの性能改善が顕著であった (誤り削減率は 35.13%)．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たからと考えられる．一方，リコーダー，オーボエで認識率が下がった．しかし，これは 150 ~ 160 個のデータのうち，1 ~ 2 個誤認識が増えたに過ぎない．

5. おわりに

本稿では，音高による音色変化を考慮するため，F0 依存多次元正規分布を提案し，RWC 研究用音楽 DB の 19 楽器を用いた実験により，認識率の向上を確認した．今後は，この枠組みに基づいてより高性能な識別関数の設計に取り組むとともに，混合音への適用などにも取り組んでいく．

謝辞　本研究は，日本学術振興会科学研究費補助金基盤研究 (B) 第 12480090 号および 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 6) 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) 柏野，村瀬 :“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.
2) 木下 他 :“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.
3) K. D. Martin: “Sound-Source Recognition: A Theory and Computional Model”, PhD Thesis, MIT, 1999.
4) I. Fujinaga and K. MacMillan: “Realtime Recognition of Orchestral Instruments”, *Proc. of ICMC*, 2000.
5) A. Eronen: “Automatic Musical Instrument Recognition”, M.Sc. Thesis, Tampere Univ. of Tech., 2001.
6) 後藤 他 :“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.