音響的特徴に基づく楽器の階層表現の獲得と それに基づくカテゴリーレベルの楽器音認識の検討

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，音響的特徴から得られる楽器の階層表現に基づいた，未知の楽器（学習データに含まれない楽器）のカテゴリーレベルの認識について述べる．未知の楽器をどのように扱うかという問題は，楽器音認識において不可避な問題であるにも関わらず，これまでの楽器音認識の研究では，扱われてこなかった．本研究では，未知の楽器をカテゴリーレベルで認識することを提案する．まず，未知の楽器のカテゴリーレベル認識に適した楽器の階層表現を自動的に獲得する手法について述べ，この手法に基づいて得られた楽器の階層表現を用いて，未知の楽器のカテゴリーレベルの認識を行う．さらに，楽器音が既知か未知か（すなわち，学習データに含まれる楽器か否か）を判定する処理を導入することで，既知の楽器は楽器名レベルで，未知の楽器はカテゴリーレベルで認識することを実現する．実験の結果，平均約 77%の未知の楽器音をカテゴリーレベルで認識することができた．

Acoustic-feature-based Musical Instrument Hierarchy and Its Application to Category-level Musical Instrument Recognition

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 category-level recognition of unknown musical instruments (*i.e.*, musical instruments which are not contained in training data) based on an acoustic-feature-based musical instrument hierarchy. The problem of how the unknown musical instruments should be dealt with is essential in musical instrument recognition. However, this problem has not been dealt with in previous studies. In this study, we propose category-level recognition of the unknown musical instruments. First, we present a method for automatic acquisition of musical instrument hierarchy, and then, using this hierarchy, we conduct experiments on recognizing the unknown musical instruments at category level. In addition, by introducing a process determining whether musical instruments are known or unknown (*i.e.*, whether the musical instruments are contained in training data or not), we realize flexible musical recognition, that is, individual-instrument-level recognition for known musical instruments and category-level recognition for unknown musical instruments. Experimental results showed that around 77% of unknown musical instrument sounds, in average, were correctly recognized at category level.

1. はじめに

ディジタル音楽配信の普及などにより急激に増加する音楽データ（本稿では音楽音響信号を指す）から効率的に自分の欲しい曲を探し出すには，音楽を対象とした計算機による情報検索（音楽情報検索）が不可欠である．音楽情報検索を効率的に行うには，音楽データに**メタデータ**（**タグ**）を付与することが有効である．これにより，検索時間が短縮されるだけでなく，タグづけに人が適度に介入することで，自動的に抽出可能な情報と少なくとも現状では自動抽出の難しい情報（アーティスト名や録音時期など）とを統一的に扱うことができるようになるからである．実際，音楽を含むマルチメディアコンテンツに対してこのようなメタデータを付与するための標準規格 **MPEG-7**[1),2)]が制定され，さまざまな研究がなされている[3)~8)]．

MPEG-7では，タグの表記方法のみ規定されており，タグの自動付与法や利用法は一切規定されていない．これは,「最大限の利用可能性のために最小限を定める」[2)]という原則に基づくものである．しかし，抽出手法に依存するような低レベルの特徴を記述する際には問題が生じる場合がある．たとえば，楽器音の特徴を記述するために用意されているタグは，LogAttackTime, TemporalCentroid, SpectralCentroid などの低レベルのものがほとんどである．これらは，特徴抽出手法によって値が変化する．そのため，たとえば「ピアノソナタの曲を検索したい」といった場面で，これらのタグを適切に利用することが困難である．

そこで我々は，特徴抽出手法に依存しない高次のタグとして，**楽器タグ**を検討する．すなわち，その楽曲で使用されている楽器名と，それぞれの楽器パートが弾き始めた時刻・弾き終えた時刻をタグとして付与する．特にクラシック音楽では,「ピアノソナタ」「弦楽四重奏」などと分類されるように，どの楽器が使用されるかは楽曲を特徴づける重要な要素である．そのため，検索の場面でも楽器に関する情報は重要であると考えられる．また，楽器タグは，聴取者の感性や主観に依存しないという観点からも音楽に付与するタグとしてふさわしいといえる．さらに，楽器タグにより「フルートが弾き始めるところから聴く」といった楽器をキーとした頭出しをすることもできる．

音楽データへの楽器タグの付与で中心となる処理は，音からの楽器名の同定（**音源同定**）である．この処理の典型的なアプローチは，同定対象音から抽出した特徴量を，あらかじめ用意されたさまざまな楽器の音響信号（学習データという）から抽出した特徴量と比較することであるが，ここで問題となるのは，学習データに存在しない楽器（**未知楽器**と呼ぶ）の扱いである．実際の音楽では，オーケストラ向け楽器から民族楽器までさまざまな種類の楽器が使われ，また，シンセサイザーで製作者が独自に合成した楽音が使われることも少なくない．そのため，これらの多様な楽器音をあらかじめ学習データとして用意するのは不可能であり，未知楽器の扱いは不可避な問題である．しかし，これまでの音源同定に関する研究[9)~17)]では，こうした未知楽器の問題は扱われてこなかった．

表 1　慣習的な楽器の階層表現

大分類	中分類	小分類	主な楽器*
弦楽器	—–	打弦楽器	PF
		撥弦楽器	CG, UK, AG
		擦弦楽器	VN, VL, VC
管楽器	木管楽器	無簧楽器	PC, FL, RC
		単簧楽器	SS, AS, TS, BS, CL
		複簧楽器	OB, FG
	金管楽器	—–	TR, TB
打楽器	（省略）	（省略）	（省略）

*「主な楽器」の欄のアルファベットは，表 3 の楽器記号を表す．

本稿では，この未知楽器の問題を，楽器の階層表現に基づいてカテゴリー（楽器種の意味で用いる）レベルで認識することにより解決する．たとえば，バイオリンともビオラとも音色が異なるが，これらに似た楽器音（たとえば，両者の音をシンセサイザー上で混ぜて作った音など）に対しては，同定結果は「弦楽器」となる．これにより，未知楽器からも検索に重要な情報を得ることができるようになる．たとえばピアノソナタを検索したいときに，ある曲から「弦楽器」が検出されれば，それだけで検索対象から外すことが可能である．本稿ではさらに，認識対象の楽器音が既知の楽器か未知の楽器かを判定する処理についても検討する．これにより，既知の楽器に対しては楽器名の認識を，未知の楽器に対しては「楽器名は分からないが弦楽器である」といった認識を行うことが可能になる．

以下，まず **2.** でカテゴリーレベルの楽器音認識で用いる楽器の階層表現を定義する．次に **3.** で，その階層表現を用いた未知楽器のカテゴリーレベルの認識と，楽器音の既知か未知かの判定について検討し，既知楽器と未知楽器の両方に対応した楽器音認識を実現する．最後に，**4.** でまとめをする．

2. 音響的特徴に基づく楽器の階層表現

本章では，楽器音をカテゴリーレベルで認識するのに適した楽器の階層表現について検討する．

楽器の階層表現で最も一般的なものは，**表 1** に示した分類である[18)]．そこで，この分類がカテゴリーレベルの楽器音認識に適しているか検討する．この分類では，まず，楽器を「弦楽器」「管楽器」「打楽器」の 3

つに分類し，さらに，弦楽器を奏法から「打弦楽器」「撥弦楽器」「擦弦楽器」に，管楽器を「木管楽器」と「金管楽器」にわけ，木管楽器をリードの有無や形状から「無簧楽器」「単簧楽器」「複簧楽器」に分類する．この分類は古くから用いられてきたが，楽器の発音機構や奏法の一部にのみ着目した分類であり，音色の総合的な類似性をとらえたものではない．たとえば，ピアノ（打弦楽器）やギター（撥弦楽器）と，バイオリン（ノーマル奏法では擦弦楽器）は，共に弦楽器であるが音色は大きく異なる．このような音色の大きく異なる楽器を同一カテゴリーとして扱うと，カテゴリーの分布が広範囲に渡ってしまい，精度よく認識するのが困難になる．

そこで本研究では，音色（ここでは楽器の音響的特徴を指す）の総合的な類似性を反映した楽器の階層表現を自動的に獲得することを検討する．ただし，楽器のどの音響的特徴に着目するかは，音源同定手法（言い換えれば楽器タグ自動付与モジュール）によって異なる．そこで，音源同定手法が用いるものと同じ特徴空間を用いて楽器を階層的に分類する．これにより，異なる音源同定手法に対して，それぞれに最適な楽器の階層表現を自動的に得ることが可能になる．

2.1 楽器の階層表現の獲得手法

本研究では，楽器音の音響的特徴を表す特徴空間に基づいて楽器を階層的に分類する．これを行う 1 つの方法は，各楽器 1 つずつ用意された音響信号から楽器間の特徴空間上の距離を算出し，距離の近い楽器対を 1 つのクラスタとして順にまとめあげていくことである．これは階層的クラスタリングと呼ばれ，さまざまな分野でよく用いられる方法である．しかし，楽器の音色は同一楽器であっても音高や個体差などによって大きく変化するため，各楽器 1 つずつの音響信号から信頼ある結果を得るのは困難である．

そこで本稿では，各楽器多数の音響信号を用意し，特徴空間上で各楽器の分布を多次元正規分布で近似する．そして，多次元正規分布に対してクラスタリングを行えるよう，クラスタリング手法を拡張する．

(1) 特徴抽出

音源同定に用いるものと同じ特徴空間を用いる．本稿の実験では文献 17) で提案された音源同定手法を用いることを想定し，ここで用いられている特徴量を抽出する．具体的には，周波数重心などの 129 個の特徴量（概要を**表 2** に示す）を抽出し，主成分分析で 129 次元から 79 次元へ圧縮（累積寄与率：99%）し，さらに線形判別分析で次元圧縮を行う（圧縮後の次元数は，学習データの楽器数 − 1）．

表 2 文献 17) で用いられた 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 3 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	それ		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)	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.

(2) 楽器間の距離の算出

各楽器の音響信号が，上記の特徴空間上で多次元正規分布に従うと仮定し，各楽器ω_iの分布の平均$\boldsymbol{\mu}_i$と共分散Σ_iを求める．そして，各楽器間のマハラノビス汎距離 $D_M(\omega_i,\omega_j)$ を次式により求める：

$$D_M(\omega_i,\omega_j) = (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)'\Sigma^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j).$$

ここで「′」は転置を表し，$\Sigma = (\Sigma_i + \Sigma_j)/2$ である．

(3) 階層的クラスタリング

上記により得られた楽器間の距離を用いて階層的クラスタリングを行う．ここでは群平均法（ある対象からクラスタまでの距離を，そのクラスタの各要素までの距離の平均とする手法）を用いる．

2.2 クラスタリング実験

音響的特徴に基づく楽器の階層表現を自動的に獲得する実験を，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[19]を使って行う．このデータベースは，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，

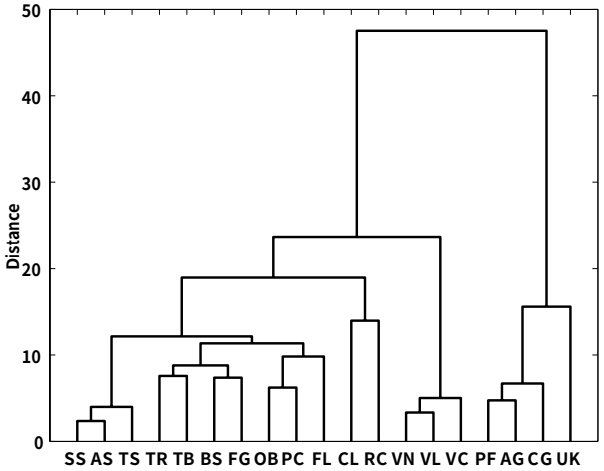


図 1　提案手法によって得られた楽器の階層表現（表 3 のすべてのデータを用いた場合）

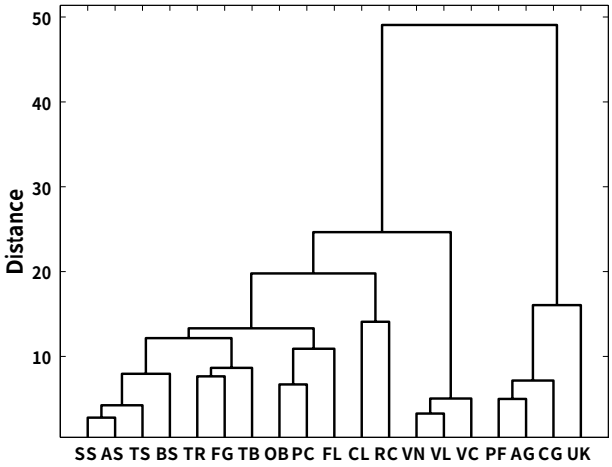


図 2　提案手法によって得られた楽器の階層表現（表 3 の半分のデータを用いた場合）

表 4　図 1 に基づいて得られた 3 つのレベルの楽器カテゴリー

大分類	中分類	小分類	属する楽器
減衰系楽器	——	ウクレレ以外	PF, CG, AG
		ウクレレ	UK
持続系楽器	弦楽器	——	VN, VL, VC
	管楽器	サックス	SS, AS, TS
		クラリネット	CL
		リコーダ	RC
		上記以外 A	TR, TB, BS, FG
		上記以外 B	OB, PC, FL

「大分類」「中分類」「小分類」は，図 1 において距離がそれぞれ 30，20，10 以下の楽器を同一クラスタとして得られた楽器カテゴリー．また，それぞれの分類名は，人間が解釈して与えている．

表 5　図 2 に基づいて得られた 3 つのレベルの楽器カテゴリー

大分類	中分類	小分類	属する楽器
減衰系楽器	——	ウクレレ以外	PF, CG, AG
		ウクレレ	UK
持続系楽器	弦楽器	——	VN, VL, VC
	管楽器	サックス	SS, AS, TS, BS
		クラリネット	CL
		リコーダ	RC
		上記以外 A	TR, TB, FG
		上記以外 B	OB, PC, FL

「大分類」「中分類」「小分類」は，図 1 において距離がそれぞれ 30，20，10 以下の楽器を同一クラスタとして得られた楽器カテゴリー．また，それぞれの分類名は人間が解釈して与えている．

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：6247 個）の内訳を**表 3** に示す．これらのデータ（各楽器約 130〜700 個程度）を使って多次元正規分布の平均と共分散を算出し，階層的クラスタリングを行う．

提案手法によって得られた楽器の階層表現を**図 1** に示す．これにより得られた階層表現に対して，距離がしきい値以下の楽器を 1 つのクラスタにまとめることで，さまざまな粒度の楽器カテゴリーを得ることができる．例として，しきい値を 10，20，30 としたときの楽器カテゴリーを**表 4** に示す．

次に，次節で行う実験の学習データ（表 3 の 6,247 音からランダムに選んだ半分のデータ）のみを使って得られた階層表現を**図 2** に示す．同様に，距離がしきい値（10，20，30）以下の楽器を 1 つのクラスタにまとめることで，**表 5** の楽器カテゴリーが得られる．

2.3　カテゴリーレベルの楽器音認識の予備実験

次章で扱う未知楽器に先立ち，既知の楽器についてカテゴリーレベルの認識精度を評価する．楽器の発音機構に基づく慣習的な楽器カテゴリー（表 1）と，提案手法によって得られた楽器カテゴリー（表 4，表 5）とを使ってカテゴリーレベルの認識率を算出（共に小分類を用いる）し，認識率の高いほうが，その音源同定手法がカテゴリーレベルのタグづけをするのに，より適した階層表現であると判断する．ただし，両者のクラス数をそろえるために，表 1 の単簧楽器をサックスとクラリネットに分け，8 クラスとして扱う．音源同定手法は文献 17) で提案したものを用い，表 3 に示す 19 楽器 6,247 音のデータに対して，ランダムに選んだ半分のデータを学習用に割り当て，残りを評価用に割り当てて認識率を求める．

実験結果を**表 6** に示す．表より，楽器の発音機構に基づく慣習的な楽器カテゴリー（表 1）を用いた場合に比べ，提案手法によって自動的に得た楽器カテゴリーを用いた場合（表 5）のほうが，カテゴリーレベルの認識率が 2.4%高かった．よって，表 5 の楽器カテゴリーは，表 1 の楽器カテゴリーよりも，文献 17) の音源同

表 6 既知の楽器に対する音源同定結果

楽器記号	楽器名レベル	カテゴリーレベル		
		Conv.	*Prop. (1)*	*Prop. (2)*
PF	80.45%	80.45%	98.12%	98.12%
CG	92.66%	96.64%	99.39%	99.39%
UK	96.73%	96.73%	96.73%	96.73%
AG	78.40%	95.73%	98.13%	98.13%
VN	71.63%	98.94%	98.94%	98.94%
VL	73.20%	92.00%	92.00%	92.00%
VC	75.18%	96.72%	96.72%	96.72%
TR	71.62%	74.32%	91.89%	82.43%
TB	74.05%	83.97%	92.37%	85.50%
SS	53.93%	78.65%	74.16%	78.65%
AS	49.17%	73.33%	69.17%	73.33%
TS	49.04%	87.50%	72.12%	87.50%
BS	67.86%	85.71%	78.57%	85.71%
OB	63.41%	70.73%	68.29%	68.29%
FG	71.23%	74.66%	75.34%	78.08%
CL	90.98%	90.98%	90.98%	90.98%
PC	80.74%	88.99%	88.99%	88.99%
FL	63.63%	66.23%	70.13%	70.13%
RC	88.88%	88.88%	88.88%	88.88%
平均	75.98%	88.85%	90.81%	91.25%

Conv.：楽器カテゴリーとして慣習的な分類（表 1）を用いた場合.
Prop. (1)：楽器カテゴリーとして提案手法によって自動的に得た分類（表 4 の小分類）を用いた場合.
Prop. (2)：楽器カテゴリーとして提案手法によって自動的に得た分類（表 5 の小分類）を用いた場合.

定手法による認識に適しているといえる.

2.4 考　　察

本研究で得られた楽器の階層表現について考察する.

- 本研究で得られた階層表現では，楽器を「減衰系楽器」と「持続系楽器」とに分け，持続系楽器を弦楽器と管楽器に分類している．これは，文献 12), 15), 16) でも用いられているものと等しい（ただし，これらの文献では階層表現を人手で与えている）．このことは，本研究で得られた階層表現が，人が経験的に得たものと一致していることを示している.
- 表 4 の小分類では，クラリネットとリコーダがそれぞれ単独で 1 つのカテゴリーになっている．これは，クラリネットに関しては「偶数次倍音（特に 2 次倍音）が弱い」というクラリネット特有の特徴[20]によるものと考えられる．リコーダに関しては，ナイフエッジに誰でも正しく呼気を吹き当てられるように導路を設けている[18]ため，吹き方などによる特徴変動があまりなかったからと考えられる．実際，リコーダの分布の分散は，19 楽器のなかで最も小さかった．そのため，リコーダと他の楽器とのマハラノビス距離が大きくなったと考えられる.
- ウクレレ・クラリネット・リコーダなど，楽器名レベルの認識率の高い楽器は，表 4・表 5 の小分類において，単独でカテゴリーを形成する傾向が見られた．これは，楽器名レベルの認識率の高い楽器は，他の楽器の分布からのマハラノビス汎距離が大きいためである．そのため，楽器名レベルで精度よく認識可能な楽器は，カテゴリーレベルで認識しても楽器名レベルと同じ情報を得ることができる.

表 7 未知楽器の認識実験に用いた楽器音

楽器の種類（楽器記号）	エレクトリックピアノ (ElecPf), シンセストリングス (SynStr), シンセブラス (SynBrs)
バリエーション	各楽器 2 種類ずつ
ベロシティ	100
音　　域	C3–C5 (A4=440Hz)

2.5 関連研究との比較

本研究で扱った，楽器の階層表現を自動的に獲得するという問題は，これまでの研究では扱われてこなかった.

聴覚心理学の研究分野では，人間の知覚の観点から楽器音の類似性についてさまざまな聴取実験が行われてきた[22)~24)]．人間の知覚に基づく楽器の階層表現は，種々の応用に有効と考えられるが，本研究の目的には必ずしも有効ではない．計算機が楽器音をカテゴリーレベルで認識する場合，計算機からみた楽器の音響的類似性を反映した階層表現を用いることが重要であり，それは，人間の知覚に基づくものとは異なるからである.

Martin[12)]，Eronen ら[15)]，Peeters ら[16)]は，音源同定を階層的に行っている．ここで用いられている階層表現は本研究のものと一部一致するが，人手で与えており，音源同定に最適な階層表現を自動的に獲得するという視点には至っていない.

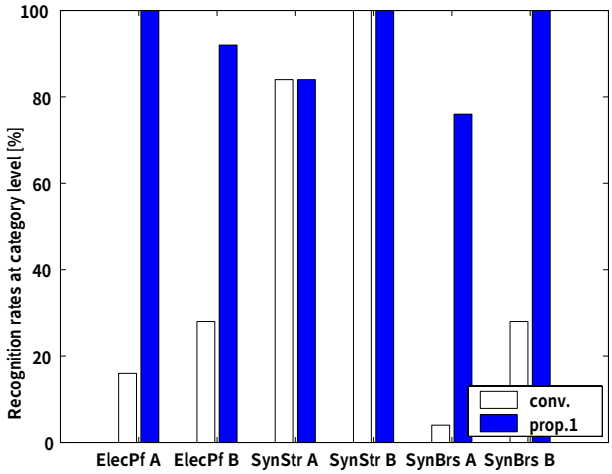
Casey[7)]は，楽器を含む音一般に対して，さまざまな音の関係を木構造により記述できることを指摘し，そのために用意された MPEG-7 の枠組みについて述べている．しかし，こうした階層表現を記述する枠組みの紹介にとどまり，階層表現を自動的に獲得する問題は扱っていない.

Dubnov ら[25)]は，電子楽器から用意した 31 個の音響信号を使って階層的クラスタリングを行っている．楽器音の音響的特徴から階層表現を得るという点は本研究と共通であるが，各楽器 1 つずつの音響信号しか使っておらず，音高や楽器の個体差などによる特徴変動は考慮されていない.

3. 未知楽器のカテゴリーレベルの認識

本章では，**2.** で得られた楽器の階層表現に基づいて，未知楽器音をカテゴリーレベルで認識する実験を行う．ここでは，未知楽器音として市販の MIDI 音源 MU2000（ヤマハ製）に収録されている電子楽器音（**表 7**）を用いる．これらの電子楽器音を，表 3 の実楽器音で学習した音源同定モジュールを使って認識する.

ここであげた電子楽器音は，実在する楽器ではない



conv.：楽器カテゴリーとして慣習的な分類（表 1）を用いた場合．
prop.1：楽器カテゴリーとして提案手法によって自動的に得た分類（表 4）を用いた場合．

図 3　未知楽器のカテゴリーレベルの認識実験の結果

ため楽器名の同定はできないが，カテゴリーレベルでは認識できることが望ましい．たとえば，ピアノ（実楽器）を知っている人が初めてエレクトリックピアノ（電子楽器）の音を聞いた場合，ピアノとは異なる音色であるために楽器名の同定はできないが，カテゴリーとしては，ピアノと同じカテゴリーの楽器と判断すると考えられる．本研究の狙いは，このような判断を計算機上で実現させることにある．

以下，まず，表 7 に示す電子楽器音をカテゴリーレベルで認識する実験を行う．次に，楽器音が既知の楽器か未知の楽器かを判定する処理について検討する．最後に，これらを組み合わせることで，既知の楽器なら楽器名を，未知の楽器ならカテゴリー名を出力する実験について述べる．

3.1　未知楽器のカテゴリーレベルの認識

表 3 のデータをすべて学習用に割り当て，表 7 のデータに対して，カテゴリーレベルの認識実験を行う．楽器カテゴリーとして，楽器の発音機構に基づく慣習なもの（表 1）と，提案手法によって自動的に得られたもの（表 4 の小分類）とを用いる．

実験結果を**図 3** に示す．提案手法によって得られた楽器カテゴリー（表 4）を用いた場合，75〜100%の割合で正しく認識が行われた．一方，慣習的な楽器カテゴリー（表 1）を用いた場合では，シンセストリングスの認識率は変わらなかったが，他の楽器の認識率は極めて低かった．このことは，楽器の発音機構に基づく慣習的な楽器分類は，機械的な発音機構の持たない電子楽器音には有効でない可能性を示唆している．

3.2　既知の楽器か未知の楽器かの判定

未知の楽器に対してはカテゴリーレベルの認識は有用であるが，既知の楽器に対しては楽器名レベルで認識するのが望ましい．これを実現するには，認識対象の楽器音が「既知」か「未知」かを判定して，認識結果を切り替える必要がある．これは，楽器名レベルの認識結果に対するリジェクションに相当し，次の手法で行う：

(1)　認識対象音を楽器名レベルで認識する．
(2)　認識対象音から，(1) で出力された楽器名の特徴空間上の分布までのマハラノビス距離を算出する．
(3)　このマハラノビス距離がしきい値以下なら「既知」，しきい値以上なら「未知」と判定する．

なお，マハラノビス距離を算出するのに使用する特徴空間は，(1) の楽器名レベルの認識で用いるものとは異なっていてもよい．

3.2.1　実験条件

表 3 の 19 楽器 6,247 音からランダムに選んだ半分のデータを学習用に割り当て，残りを既知楽器の認識用，表 7 のデータを未知楽器の認識用に割り当てる．特徴空間は，表 2 の 129 個の特徴量を抽出して得られる特徴空間を，主成分分析のみで 23 次元（累積寄与率：90%）に圧縮したもの，主成分分析のみで 18 次元（累積寄与率：88%）に圧縮したもの，主成分分析で 79 次元（累積寄与率：99%）に圧縮後線形判別分析で 18 次元に圧縮したものをそれぞれ使用し，結果を比較する．なお，この実験では，既知楽器で，最初の楽器名レベルの認識で誤ったデータは，評価対象から除外する．

3.2.2　実験結果

実験結果を**表 8** に示す．主成分分析による 23 次元特徴空間を用いてしきい値を 40 としたとき，約 85%の割合で正しく既知楽器と未知楽器の判定を行うことができた．「既知楽器を正しく既知と判定する割合」と「未知楽器を正しく未知と判定する割合」はトレードオフの関係にあるが，両者の平均をとると 23 次元特徴空間ではおおよそ 80〜85%程度であった．

「既知楽器を正しく既知と判定した割合」が 85〜86%のときに着目すると，「未知楽器を正しく未知と判定した割合」は，主成分分析による 23 次元特徴空間を用いた場合が最も高く 85%で，主成分分析と線形判別分析を併用した場合が 71%と最も低かった．線形判別分析は，学習データの分布間の分離（クラス内分散・クラス間分散比）を最大にする次元圧縮法で，楽器名の同定には有効であることが示されている[17]．しかし，クラス内分散・クラス間分散比最大化に寄与する特徴量が，既知か未知かの判定に有効とは限らない．そのため，線形判別分析で得られた特徴空間は，既知か未知かの判定には適していなかったと考えられる．

楽器別に見ると，ElecPf A を既知楽器と誤判定することが多かった．これは，この音が比較的本物のピア

表 8 既知楽器か未知楽器かの判定（未知楽器のリジェクション）に関する実験結果

	次元圧縮*	PCA(23)				PCA(18)			PCA+LDA(18)		
	しきい値	50	40	30	25	40	30	25	40	30	25
既知楽器	PF	92%	86%	79%	71%	93%	84%	79%	88%	82%	71%
	CG	94%	90%	83%	77%	95%	89%	86%	97%	92%	85%
	UK	86%	82%	68%	63%	87%	81%	73%	88%	82%	73%
	AG	91%	86%	80%	75%	90%	83%	80%	92%	86%	78%
	VN	91%	86%	73%	61%	94%	85%	76%	94%	84%	73%
	VL	95%	94%	79%	70%	97%	95%	85%	97%	93%	86%
	VC	96%	93%	89%	79%	97%	93%	92%	99%	94%	87%
	TR	94%	87%	70%	60%	96%	89%	79%	96%	92%	50%
	TB	92%	86%	75%	66%	95%	89%	84%	97%	91%	84%
	SS	96%	88%	73%	54%	96%	85%	71%	96%	94%	85%
	AS	88%	81%	58%	50%	92%	86%	76%	88%	80%	71%
	TS	80%	62%	46%	34%	80%	78%	70%	88%	70%	58%
	BS	88%	73%	63%	51%	92%	77%	69%	88%	77%	82%
	OB	87%	75%	65%	54%	87%	79%	71%	98%	85%	61%
	FG	85%	78%	68%	64%	87%	78%	74%	89%	78%	67%
	CL	92%	77%	67%	52%	90%	85%	80%	98%	90%	76%
	PC	90%	82%	67%	55%	92%	83%	77%	82%	73%	35%
	FL	88%	71%	47%	37%	96%	80%	50%	100%	88%	40%
	RC	91%	81%	69%	53%	94%	81%	72%	95%	90%	59%
	平均	91%	85%	74%	65%	93%	86%	79%	94%	86%	72%
未知楽器	ElecPf A	36%	44%	64%	76%	32%	36%	36%	24%	44%	48%
	ElecPf B	52%	84%	88%	92%	36&	52%	55%	36%	60%	76%
	SynStr A	100%	100%	100%	100%	100%	100%	100%	56%	88%	92%
	SynStr B	100%	100%	100%	100%	100%	100%	100%	40%	60%	100%
	SynBrs A	76%	80%	88%	92%	72%	84%	88%	72%	80%	84%
	SynBrs B	100%	100%	100%	100%	100%	100%	100%	76%	96%	100%
	平均	77%	85%	90%	93%	73%	79%	81%	51%	71%	83%

*「次元圧縮」欄の PCA は主成分分析のみで次元圧縮，PCA+LDA は主成分分析で次元圧縮した後，線形判別分析でさらに次元圧縮したことを示す．PCA，PCA+LDA の後のかっこ内の数字は，圧縮後の次元数を表す．

表 9 既知楽器／未知楽器両方に対する楽器音認識実験

	正解 (a)	正解 (b)	不正解
PF	68.80%	17.29%	13.91%
CG	83.49%	11.62%	4.89%
UK	96.73%	—–	3.27%
AG	68.27%	14.40%	17.33%
VN	61.70%	13.82%	24.48%
VL	68.80%	11.20%	20.00%
VC	69.71%	9.85%	20.44%
TR	63.51%	14.86%	21.63%
TB	63.36%	16.79%	19.85%
SS	47.19%	11.24%	41.57%
AS	40.00%	16.67%	43.33%
TS	29.81%	25.96%	44.23%
BS	49.11%	19.64%	31.25%
OB	47.56%	19.51%	32.93%
FG	56.16%	16.44%	27.40%
CL	90.98%	—–	9.02%
PC	66.06%	17.43%	16.51%
FL	45.45%	19.48%	35.07%
RC	88.88%	—–	11.12%
平均	66.62%	13.16%	20.22%
ElecPf A	—–	44.00%	56.00%
ElecPf B	—–	76.00%	24.00%
SynStr A	—–	88.00%	12.00%
SynStr B	—–	100.00%	0.00%
SynBrs A	—–	60.00%	40.00%
SynBrs B	—–	96.00%	4.00%
平均	—–	77.33%	22.67%

正解 (a)：楽器名レベルで正解．
正解 (b)：楽器名レベルの認識結果を棄却し，カテゴリーレベルで正解．

ノの音に近かったからと考えられる．実際，ElecPf A は，人間が聞いても，本物のピアノの音をフィルタ処理により音質を変えたようにも聞こえる．今後は，どの程度既知楽器と音色が違えば未知楽器として扱うべきか，人間の場合と比較しながら適切に決めていくことも必要である．

3.3 既知楽器には楽器名を，未知楽器にはカテゴリー名を出力する楽器音認識

以上の処理を組み合わせて，既知の楽器なら楽器名を，未知の楽器ならカテゴリー名を出力する実験を行う．前節の実験と同様に，表 3 の 19 楽器 6,247 音からランダムに選んだ半分のデータを学習用に割り当て，残りを既知楽器の認識用，表 7 のデータを未知楽器の認識用に割り当てる．特徴空間は，表 2 の 129 個の特徴量からなる特徴空間を主成分分析で 23 次元に圧縮したものを用いる．しきい値は 40 とする．

実験結果を**表 9** に示す．既知楽器に関しては，66.62%の楽器音に対して楽器名レベルで正しく認識することができ，13.16%の楽器音に対して楽器名レベルの認識結果を棄却したものの，カテゴリーレベルで正しく認識することができた．未知楽器に関しては，77.33%の楽器音に対して，楽器名レベルの認識結果を棄却した上で，正しくカテゴリーレベルで認識することができた．その結果，既知楽器の平均認識誤り率は 20.22%，未知楽器の平均認識誤り率は 22.67%となった．

3.4 考　　察

本実験に対する考察を以下にまとめる：

- 本研究では，未知楽器に対して「楽器名はわからないが弦楽器」といった認識アプローチを可能にした．このアプローチは，多様な楽器音に対して適切なタグづけを可能にするだけでなく，自動採譜にも有用である．たとえば，ピアノの音と，ピアノとは異なるが似ている未知の音（エレクトリックピアノ）が共に含まれる音楽を採譜するとき，従来の楽器音認識では区別ができなかった．本研究では，未知楽器をカテゴリーレベルで認識することで，このような区別を可能にした．
- 本研究のアプローチは，他のメディアとの情報統合にも有用であると考えられる．たとえば，音楽演奏の映像に対するタグづけを考えたとき，音から「楽器名はわからないが弦楽器」と認識された楽器に対して，画像からある民族楽器であることがわかれば，弦楽器に属する新たな楽器として再学習する，といった処理にも応用できる．
- 本実験では，楽器カテゴリーとしてすべて小分類

を用いた．これは，多くの応用においては，大分類・中分類は粒度が大きすぎると判断したからである．しかし，本来，どの程度細かいカテゴリーを用いるべきかは応用に依存する．そのため，楽器の階層表現から楽器カテゴリーを得る際のしきい値を，応用に応じて適切に（できれば自動的に）決めることが必要になる．

- 本研究では，楽器音認識に適した楽器の階層表現を得たが，人間が検索クエリーとして楽器カテゴリーを指定する場合には，人間の直感に合った楽器の階層表現が必要である．そこで，今後は，人間の直感に合った楽器の階層表現を聴覚心理学の実験結果[22)~24)]に基づいて構築し，2つの楽器の階層表現間を変換することで，楽器による音楽検索を実現する．このような，同一概念を表した異なる複数の階層表現間の変換問題は，**ontology problem** あるいは **problem of semantic mapping** と呼ばれ[26)]，オントロジー工学における共通の問題の1つである．

4. おわりに

本稿では，音楽データへ楽器タグを付与する際に問題となる事柄として，学習データに存在しない楽器（未知楽器）をどう扱うかという問題をあげ，これをカテゴリーレベルで認識することを提案した．これは，人が初めて聞いた音に対して感じるような「聞いたことのない音だけど，弦楽器系の音だと思う」という判断を，計算機上で実現するアプローチである．実楽器音で学習した音源同定モジュールに対して未知の電子楽器音を入力させたところ，約77%の未知楽器音の楽器カテゴリーを正しく認識することができた．

今後は，混合音や実際の楽曲に対して，提案手法の有効性を確認していく予定である．

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号およびサウンド技術振興財団研究助成による．また，本研究の実験において，文献19)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，有益なご助言をくださった片寄 晴弘氏（関西学院大学），柏野 邦夫氏（NTTコミュニケーション科学基礎研究所），中臺 一博氏（株式会社ホンダ・リサーチ・インスティチュート・ジャパン）に感謝する．

参考文献

1) http://ipsi.fhg.de/delite/Projects/MPEG7/
2) http://www.itscj.ipsj.or.jp/mpeg7/
3) S. Chang, A. Puri, T. Sikora and H. Zhang (ed.): The Special Issue on MPEG-7, *IEEE Trans. Circuits Syst. Video Technol.*, **11**, 6, pp.685–772, 2001.
4) P. Herrera and X. Serra: A Proposal for the Description of Audio in the Context of MPEG-7, *Proc. European Workshop on Content-based Multimedia Indexing*, 1999.
5) G. Peeters, S. McAdams and P. Herrera: Instrument Sound Description in the Context of MPEG-7, *Proc. ICMC*, 2000.
6) J. Hunter: Adding Multimedia to the Semantic Web—Building an MPEG-7 Ontology, *Int'l Semantic Web Working Sympo.*, 2001.
7) M. Casey: General Sound Classification and Similarity in MPEG-7, *Organaized Sound*, **6**, 2, 2002.
8) E. Gomez, F. Gouyon, P. Herrera and X. Amatriain: Using and Enhancing the Current MPEG-7 Standard for a Music Content Processing Tool, *Audio Engineering Society 114th Convention*, 2003.
9) 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.
10) 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.
11) 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.
12) K. D. Martin: Sound-Source Recognition: A Theory and Computional Model, PhD Thesis, MIT, 1999.
13) J. C. Brown: Computer Identification of Musical Instruments using Pattern Recognition with Cepstral Coefficients as Features, *J. Acoust. Soc. Am.*, **103**, 3, pp.1933–1941, 1999.
14) I. Fujinaga and K. MacMillan: Realtime Recognition of Orchestral Instruments, *Proc. ICMC*, 2000.
15) A. Eronen and A. Klapuri: Musical Instrument Recognition using Cepstral Coefficients and Temporal Features, *Proc. ICASSP*, pp.753–756, 2000.
16) G. Peeters and X. Rodet: Automatically Selecting Signal Descriptors for Sound Classification, *Proc. ICMC*, 2002.
17) 北原 鉄朗，後藤 真孝，奥乃 博：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，情処研報，2002-MUS-46, pp.1–8, 2002.
18) 早坂 寿雄：楽器の科学，電子情報通信学会, 1992.
19) 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.
20) N. F. Fletcher and T. D. Rossing: 楽器の物理学，シュプリンガー・フェアラーク東京，2002.
21) 安藤 由典：楽器の音響学，音楽之友社，1996.
22) L. Wedin and G. Goude: Dimension Analysis of the Perception of Instrument Timbre, *Scand. J. Psychol.*, **13**, pp.228–240, 1972.
23) J. M. Grey: Multidimensional Perceptual Scaling of Musical Timbres, *J. Acoust. Soc. Am.*. **61**, 5, pp.1270–1277, 1977.
24) P. Toiviainen, K. Kaipainen and J. Louhivuori: Musical Timbre: Similarity Ratings Correlate with Computational Feature Space Distances, *J. New Music Research*, **24**, pp.292–298, 1995.
25) S. Dubnov and N. Tishby: Clustering of Musical Sounds using Polyspectral Distance Measures, *Proc. ICMC*, 1995.
26) F. Nack and L. Hardman: Towards a Syntax for Multimedia Semantics, *CWI Reports of INS 2 (Multimedia and Human-Computer Interaction)*, INS-R0204, 2002.