音 楽 情 報 科 学　40−2
（２００１．５．２３）

音高による音色変化に着目した音源同定手法

北原 鉄朗[†]　　後藤 真孝[††]　　奥乃 博[†††]

[†] 東京理科大学理工学部情報科学科
[††] 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
[†††] 京都大学大学院情報学研究科知能情報学専攻知能メディア講座

tetsu@cs.is.noda.sut.ac.jp　　m.goto@aist.go.jp　　okuno@nue.org

あらまし　本稿では, 音高によって音色が変化することに着目した, 新たな音源同定手法を提案する. 我々は以前, 同一楽器であっても音高によって特徴量が変化することを指摘し, 特徴量テンプレート (多数の特徴量の集合) を音高によって使い分けることを提案した. 本稿では, この性質により積極的に対応するために, 各特徴量を音高による変化の仕方で 3 種類 (無相関型特徴量・連続変化型特徴量・離散変化型特徴量) に分類する. その上で, 各特徴量の音高による変化の仕方を表現するための, 入力信号の音高に関する関数を提案する. また, 楽器の階層的分類を考慮し, 階層的処理を導入する. 本手法を実装・実験した結果, 提案手法の有効性を確かめることができた.

Musical instrument identification considering pitch-dependent characteristics of timbre

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

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

Abstract　This paper describes a new musical instrument identification method considering pitch-dependent characteristics of timbre. In our previous paper, we proposed a method of selecting feature templetes (i.e., sets of features) according to the pitch because timbre depends on the pitch. In this paper, we extended the method to classify features into three types (non-correlative features, continous-change features and discrete-change features), to calculate representative values and measures of dispersion according to the type, and to consider the hierarchy of instruments. Experimental results showed that the performance of musical instrument identification is improved by the proposed method.

1. はじめに

音楽情景分析[1),2)] において, 音源同定は重要な問題である. 音源同定における問題点として, 次の 2 つが挙げられる:

問題点 1 (音色変化の問題)　実楽器は, 同一楽器であっても, 音高 (基本周波数)・音の強さ・楽器の個体差・演奏方法によって音色が変化するため, 同定が困難である.

問題点 2 (未知楽器の問題)　対象でない楽器 (未知の楽器) が入力されたときに, まったく同定できない.

音源同定を扱った研究として, 柏野らの OPTIMA[1),2)] がある. OPTIMA の音源同定部では, 音楽音響の分野の知見や楽器の構造などを考慮して選ばれた 41 個の特徴量を抽出し, 主成分分析・判別分析を行う. しかし, 上で挙げた 2 つの問題点は考慮されていない. 木下らの手法[3)] は, 混合音における周波数成分の重複に対する虚弱性改善を行うものである. そのため, 単音入力に対しては OPTIMA と同等である. また, 適応型混合テンプレート法[4)] は, 特徴量の抽出は行わず, 波形テンプレートのフィルタリングにより同一楽器の個体差を吸収し, 波形レベルでの相関を求めている. これは,

音色変化の問題の一部に対応したものであるが, 音高による音色変化の問題は扱っていなかった. Martin[5] は, 音色変化の問題の指摘はしているものの, 明示的に対処していない. また, 階層的な処理は導入したが, 未知楽器の問題は扱っていなかった.

我々は, 音色変化の問題に対して, 同一楽器であっても音高によって特徴量が変化することを指摘し, 入力信号の音高によって特徴量テンプレート (多数の特徴量の集合) を使い分けることを提案した[6]. そこで実装したシステムは, 1 オクターブごとの粗い音域ごとに特徴量テンプレートを作成し, 入力信号の基本周波数が属する音域の特徴量テンプレートを選択するものであった. しかし, 音域を 1 オクターブ単位で分割する根拠は明確ではなかった. また, MIDI 音源に対して有効性を確認しているものの, 実楽器の音響信号に対して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の問題の解決策として, 音高による変化の仕方で特徴量を 3 つに分類し, それらの特徴量の変化の様子を基本周波数の関数で表現する. さらに, 未知楽器の問題の解決策として, 楽器の階層的な分類に基づいて一定の階層まで同定する. たとえば, 未知の classical guitar が入力されたときには「減衰系楽器」であることまで同定する.

以下, まず **2.** で対象とする楽器とその階層的な分類を定め, 各階層で有効な特徴量を提案する. 次に **3.** で, 「音高による音色変化」に着目したときに考慮すべき課題を挙げ, その解決策として, 上で提案した「特徴量の分類」と「特徴量の変化を表現する基本周波数の関数」の詳細を述べる. **4.** で実装した音源同定システムの処理の流れについて説明し, **5.** と **6.** で評価実験について述べる. 最後に, **7.** で結論と今後の課題を述べる.

2. 楽器の分類と階層に応じた特徴量

楽器音の同定あるいは認知や理解は, 理解のレベルが多段階であるので, その表現は階層的でなければならない. 例えば, 個別の楽器, 弦楽器パートや管楽器パート, ボーカルと伴奏など, どのレベルに注目するかは聞き手にとって主体的であり, それを表現するために階層的表現が不可欠である. しかし, 楽器音の階層的表現は, 従来の研究ではあまり扱われてこなかった[5),7)].

楽器音の階層的な表現として, 本研究では楽器の発音機構に基づいた楽器の分類を使用し, 各階層に適した特徴量を提案する. このような階層的表現により, 未知楽器の問題も, 「弦楽器」「管楽器」といったより一般的な階層での同定として, 対応することができる. また, 本稿における階層的表現を他の楽音表現との相互流通のためのオントロジーとして展開していくこともできる.

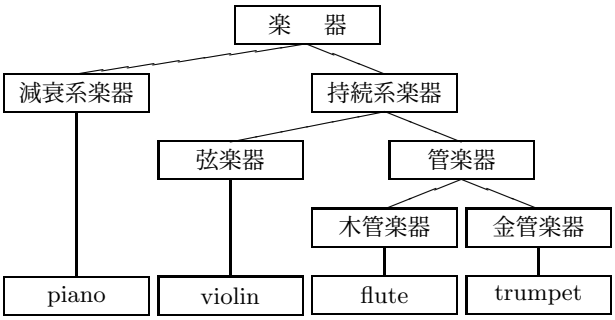


図 1　対象楽器の階層的分類

以下, **2.1** で対象楽器を階層的に分類し, **2.2** から **2.4** で各階層に適した特徴量を提案する. なお, これらの特徴量は文献 5), 8)~10) を参考にしながら決定した.

2.1 楽器の分類

対象楽器を piano, violin, flute, trumpet の 4 つとする. これら 4 つの楽器の階層的な分類を**図 1** に示す. まず, 音量の時間変化の仕方から「持続系楽器」と「減衰系楽器」に分類できる. 「持続系楽器」は, 演奏者の意志で同じ音量を持続させることができ, 音を出している最中に, (例えば弦や息の強さで) 変化を加えることができる楽器である. 一方, 「減衰系楽器」は, 発音した瞬間に音量が決まり, その後音量が小さくなっていく楽器である. 「持続系楽器」として, 本稿では「弦楽器」と「管楽器」を取り上げる. ただし, 「弦楽器」は最も代表的な弓で弦を擦る奏法のみに対応し, ピチカート奏法は対象外とする. また, すべての楽器にわたってスタッカート奏法は対象外とする☆.

2.2 第 1 階層 (減衰系楽器—持続系楽器)

piano のような「減衰系楽器」では, 定常状態がなく, 立ち上がり後はすぐ減衰に移ることが特徴である[8]. そこで, 特徴量を次の 3 つとする:

i) パワー包絡線☆☆の線形最小二乗法による近似直線の傾き.

ii) 最初から 800ms までのパワー包絡線の微分係数の中央値.

iii) 最大パワー値と 800ms のときのパワー値との差.

2.3 第 2 階層 (弦楽器—管楽器)

violin のような「弦楽器」は, 高調波成分がかなり次数の高いものまで含まれており[9], 各高調波成分のパワーの変動が激しい. そこで, 特徴量を次の 3 つとする:

i) 全持続時間の 70%以上鳴り続けている高調波成分の個数.

ii) 周波数重心 (高調波成分のパワー値を重みとした

☆ ピチカート奏法は「持続系楽器」として扱うのは困難であり, スタッカート奏法は, 音量が変化する前に発音が中断されるため, 「持続系楽器」と「減衰系楽器」の区別が困難だからである. このような多様な奏法を扱える手法の実現は, 今後の課題である.

☆☆ 本稿で単に「パワー」あるいは「振幅」と言った場合は, 各高調波成分ではなく全体のパワーを指す.

周波数値の重み付き平均. ただし, パワー値と周波数値は時間方向の中央値を用いる).

iii) 各高調波成分のパワー値の時間変化の標準偏差を全高調波成分にわたって平均した値.

2.4 第 3 階層 (木管楽器—金管楽器)

この階層では, 上記 2 つの階層に比べて同定すべき 2 つのカテゴリー間の音色が似ているため, 同定に効果的な特徴量のみを抽出するのは難しい. そこで, 識別に効果的と予想される特徴量を数多く抽出し, 主成分分析によって次元を圧縮する.

実際に用いた特徴量は, 次の 17 個である:

- スペクトルに関する特徴量

i) 周波数重心.

ii) 全体の音量に対する基音の音量の割合.

iii) 全体の音量に対する基音から 5 次までの倍音の音量の割合.

iv) 全体の音量に対する 19 次以上の倍音の音量の割合.

v) 全体の音量に対する 16 次以上の倍音の音量の割合.

vi) 基音から 6 次までの倍音と 7 次以上の倍音との音量の比.

vii) 全持続時間の 70%以上鳴り続けている高調波成分の個数.

- 立ち上がり部分の時間変化に関する特徴量

viii) 最初から 800ms までの各高調波成分のパワーの時間変化の標準偏差の和.

ix) 最初から 800ms までの基本周波数の時間変化の標準偏差.

- パワー・周波数の時間変化・変調に関する特徴量

x) 周波数重心の時間変化 (各時刻における, 高調波成分のパワー値を重みとした周波数値の重み付き平均) の標準偏差.

xi) 周波数重心の時間変化に関する分布のヒストグラム (階級の個数が 10 個になるように階級の幅を設定) における, 度数の合計に対する最大の度数の割合.

xii) 上のヒストグラムで, 相対度数が 1 割以上の階級の個数.

xiii) 振幅変調の振幅.

xiv) 振幅変調の振動数.

xv) 周波数変調の振幅.

xvi) 周波数変調の振動数.

xvii)基音のパワーの時間変化と周波数の時間変化との相関.

ここで, 振幅変調と周波数変調の振動数は, 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 (Savitzky と Golay の 2 次多項式適合による平滑化[11] を使用) と元の信号との差に対する四分位幅☆からそれぞれ算出する.

☆ 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.

表 1 各主成分の重み値の一部

主成分		重み値の大きい特徴量	重み値
第 1 (34.7%)	i)	周波数重心	0.3872
	vii)	全継続時間の 70%以上鳴り続けている高調波成分の個数	0.3675
	iii)	全体の音量に対する基音から 5 次までの倍音の音量の割合	-0.3694
第 2 (17.0%)	x)	周波数重心の時間変化の標準偏差	-0.5590
	ix)	最初から 800ms までの基本周波数の時間変化の標準偏差	-0.5517
	xv)	周波数変調の振幅	-0.5259
第 3 (12.5%)	v)	全体の音量に対する 16 次以上の倍音の音量の比	0.4454
	vi)	基音から 6 次までの倍音と 7 次以上の倍音との音量の比	0.4200
	iv)	全体の音量に対する 19 次以上の倍音の音量の割合	0.4084
第 4 (9.7%)	xiv)	振幅変調の振動数	-0.5701
	xvi)	周波数変調の振動数	-0.4569
	xiii)	振幅変調の振幅	-0.3884

注 「主成分」欄の括弧は寄与率を表す.

主成分分析

上で挙げた特徴量に関して主成分分析を行い, 次元を圧縮する. 累積寄与率は 80%とする. 各主成分に関して, 重み値☆☆の一部を**表 1** に示す. 表より, 第 1 主成分と第 3 主成分が「高調波成分の豊富さ」, 第 2 主成分が「周波数の変動の幅」, 第 4 主成分が「ビブラート・トレモロ」を総合的に表していることがわかる.

3. 音高による音色変化に着目した音源同定

本音源同定方式では, 各楽器名がラベル付けられたデータベース (個々のデータを学習データと呼ぶ) に基づいて音源同定を行う. 入力の音響信号 (入力データ) と学習データとの類似度を求め, 最も類似度の高い学習データのラベルを入力データの楽器名として出力する. ただし, 学習データが以下の理由により音高に依存することを考慮する:

(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[9].

(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[9].

(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.

この問題を適切に扱う 1 つの方法は, あらゆる音高に対して学習データを用意することである. しかし, そのように多量のデータを用意するのは不可能である. そこで, 有限の学習データに基づいてさまざまな音高に対処できる手法が必要である.

☆☆ ここでいう重み値は, 後に導入される重み値ではなく, 主成分分析の結果得られる重み値である.

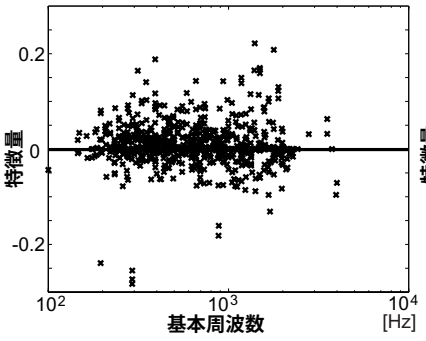


図 2 無相関型特徴量の例
(持続系楽器の ii))

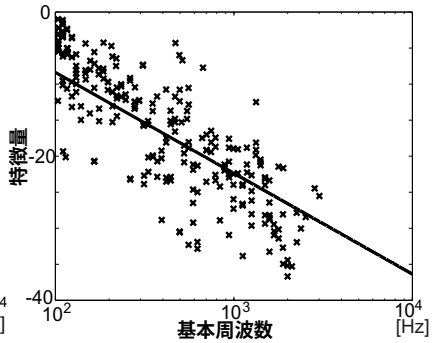


図 3 連続変化型特徴量の例
(減衰系楽器の iii))

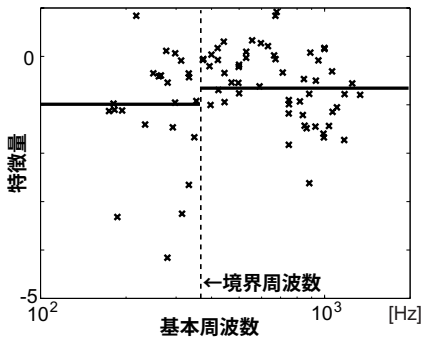


図 4 離散変化型特徴量の例
(金管楽器の第 4 主成分)

(3 つの図の直線は代表値関数を表す.)

以上から, ここで考慮すべき課題は次の 2 つにまとめられる:

課題 1 特徴量によって音高による変化の仕方が異なることをどのように考慮するか.

課題 2 有限の学習データから, 入力データの音高の特徴量をどのように推定するか.

本稿では, **課題 1** に対して, 各特徴量を音高による変化の仕方で, **無相関型特徴量**, **連続変化型特徴量**, **離散変化型特徴量**に分類する. **課題 2** に対して, 学習データの音高による変化の様子を表現するための関数として, **代表値関数**と**変動値関数** (代表値関数からの散らばりの様子を表す関数) を導入する. これらの定義は, 特徴量の 3 つの分類ごとに行う.

3.1 特徴量の 3 つの分類

(i) 無相関型特徴量

特徴量と基本周波数☆の間に, 明白な相関性が認められないとき (あるいは相関性が極めて弱いとき), その特徴量を**無相関型特徴量**と呼ぶ (**図 2**).

(ii) 連続変化型特徴量

特徴量と基本周波数の間に, ある程度以上の強さで相関性が認められるとき, その特徴量を**連続変化型特徴量**と呼ぶ (**図 3**).

(iii) 離散変化型特徴量

ある周波数が 1 つ以上存在し, その周波数前後で特徴量の分布が大きく変わるとき, その特徴量を**離散変化型特徴量**と呼び, その周波数を**境界周波数**と呼ぶ (**図 4**).

各特徴量の分類は, 自動で行うのは困難なため, 人手で行う. 離散変化型特徴量が存在する場合, 境界周波数も人手で求める. 人手により各特徴量を分類した結果を**表 2** に示す.

3.2 代表値関数

代表値関数を以下で定義する. 代表値関数は従来の手法における平均値に相当し, $\mu_s^i(f)$ で表す. ここで, i は特徴量番号, s はカテゴリー, f は入力データの基本周波数である. 周波数は対数で表す.

表 2 各特徴量の分類

第 1 階層

	減衰系楽器	持続系楽器
i)	連続変化型	無相関型
ii)	連続変化型	無相関型
iii)	連続変化型	連続変化型

第 2 階層

	弦楽器	管楽器
i)	連続変化型	連続変化型
ii)	連続変化型	連続変化型
iii)	連続変化型	連続変化型

第 3 階層

主成分	木管楽器	金管楽器
第 1	連続変化型	連続変化型
第 2	連続変化型	離散変化型
第 3	無相関型	無相関型
第 4	連続変化型	離散変化型
第 5	連続変化型	無相関型
第 6	無相関型	無相関型

(i) 無相関型特徴量の場合

各学習データの特徴量の中央値. これは, f が変化しても一定の値をとる.

(ii) 連続変化型特徴量の場合

各学習データの特徴量の線形最小二乗法による近似直線.

(iii) 離散変化型特徴量の場合

境界周波数で区切られた区間ごとの各学習データの特徴量の中央値.

3.3 変動値関数

変動値関数の定義を以下で述べる. 変動値関数は従来の手法における標準偏差に相当し, $\sigma_s^i(f)$ で表す. パラメータの意味は前節と同じである. 変動値関数の定義は 3 つの分類に対して共通であるが, 離散変化型特徴量では, 境界周波数で区切られた区間ごとに, その区間に属する学習データのみ用いる.

☆ 以下, 具体的な手法の説明では, 「音高」ではなくこちらの用語を用いる.

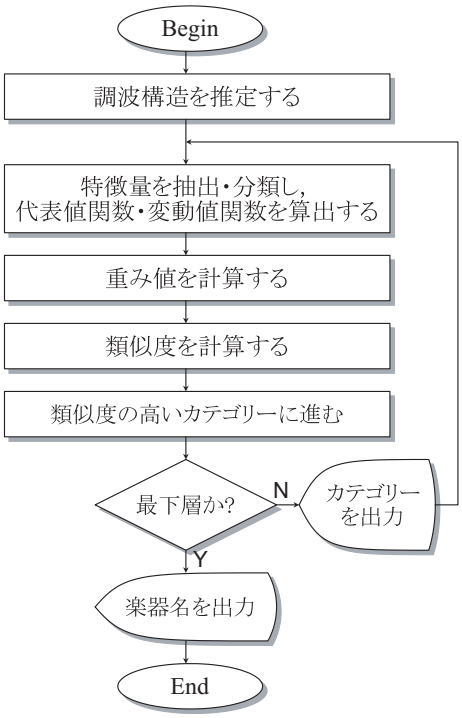


図 5　処理の概要

変動値関数は, 各学習データの特徴量と代表値関数の差の重み付き二乗平均の平方根, すなわち,

$$\sigma_k(f) = \sqrt{\sum_k w_k(f) \cdot (x_k^i - \mu_s^i(f))^2 \Big/ \sum_k w_k(f)}$$

とする. ここで, x_k^i は学習データ k から抽出された i 番目の特徴量を表す. $w_k(f)$ は, 学習データ k の重み値で次式で定義する:

$$w_k(f) = \exp(-|F_k - f|).$$

これにより, 入力データの基本周波数に近い学習データをより重視できる.

4. 処理の流れ

本章では, 提案手法に基づいて実装した音源同定システムの処理の流れを述べる. システムの実装には MATLAB を用いた.

4.1 概　　要

処理の概要を**図 5** に示す. 入力信号として楽器の単音を収録した音響信号を受け付け, それに対応する楽器名を出力する. システムには**特徴量テンプレート**を用意しておく. 特徴量テンプレートとは, 類似度計算に用いられる学習データの集合である. ここで, 特徴量テンプレート作成に使用した音響信号も, 入力信号と同じく単音を収録した音響信号とする.

図 1 に基づき, 最も抽象度の高い階層 (第 1 階層) からより低い階層へと, 反復的に入力データのカテゴリーを同定する. ここでは二分木のため, 1 回の反復処理において, 同定すべきカテゴリーの個数は 2 つとなる.

まず, 反復前の前処理として, 入力信号の調波構造を推定する. 以下, 反復処理として, (1) 特徴量の抽出・分類と代表値関数・変動値関数の算出, (2) 重み値・類似度の計算, を行う. 類似度の高いカテゴリーが最下層であれば, そのカテゴリーの楽器名を出力して処理を終了し, そうでなければ, 1 段低い階層に移り, 反復処理に戻る.

4.2 調波構造の推定

入力信号に対し, 調波構造を推定する. 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 (時間分解能: 10ms), 各時刻において, パワースペクトルの周波数方向の 3 次導関数の零交差からピーク抽出を行う☆. パワー値が最大のピークの周波数 (あるいは, その整数分の 1 の周波数のピークが存在すれば, それらのうち最大公約数) を基本周波数とし, 抽出されたピークから調波構造 (20 次倍音まで) を抽出する. また, 基本周波数の時間変化の中央値をその音響信号を代表する基本周波数とする. 周波数・パワーとも対数で表し (特にパワーは [dB]), 周波数は基本周波数が 0, パワーは最大値が 100dB となるように正規化する.

4.3 特徴量の抽出・分類と代表値関数・変動値関数の算出

2. で提案した特徴量を抽出する. 特徴量抽出後は, 平均が 0, 分散が 1 となるように正規化する.

そして, **3.** で述べたように各特徴量を分類し, 代表値関数 $\mu_s^i(f)$ と変動値関数 $\sigma_s^i(f)$ を用いる.

4.4 重み値の計算

2 つのカテゴリー s, t に対して, 各特徴量 i の重み値 $W_s^i(f), W_t^i(f)$ を計算する. 次の 2 つの条件を満たすとき, カテゴリー s にとって i 番目の特徴量は重要であると考えられる[3]:

(1) カテゴリー s, t の特徴量が大きく異なり, 2 つの代表値関数 $\mu_s^i(f)$ と $\mu_t^i(f)$ が十分に離れている.

(2) その特徴量が安定していて, 変動値関数 $\sigma_s^i(f)$ が十分に小さい.

よって, 重み値 $W_s^i(f)$ を次式により定義する:

$$W_s^i(f) = \sqrt{P\left(|X| \le \frac{|\mu_t^i(f) - \mu_s^i(f)|}{\sigma_s^i(f)}\right)^2}.$$

ここで, P は標準正規確率分布の確率値, すなわち,

$$P(|X| \le z) = \int_{-z}^{z} (1/\sqrt{2}) \exp(-x^2/2) dx$$

である ($W_t^i(f)$ も同様に定義する). ただし, 特徴量 i が

☆ 文献 12) 参照. ピーク位置は奇数次の導関数の零交差の位置から求めることができる. 導関数の次数が高くなればなるほど, 互いに接近しているピークの分離はよくなるが, ノイズに対して敏感になる.

両方のカテゴリーにわたって同一の値を示すとき, すなわち, $\sigma_s^i(f), |\mu_t^i(f) - \mu_s^i(f)|$ がともに 0 であるとき, $W_s^i(f) = 0$ とする.

4.5 類似度の計算

入力信号から抽出された特徴量 x^i と特徴量テンプレート内の学習データとの類似度を, カテゴリーごとに次式により計算する:

$$R_s(f) = \exp\left(\sum_i W_s^i(f) \cdot \log d_s^i(f) \Big/ \sum_i W_s^i(f)\right).$$

ただし,

$$d_s^i(f) = P\left(|X| \geq \frac{|x^i - \mu_s^i(f)|}{\sigma_s^i(f)}\right).$$

4.5 および **4.6** で定義した式は, 木下らの式[3](を変形したもの) に基づいているが, 入力信号の基本周波数に関する関数として定義している点で異なる. また, 文献 6) では, パラメータとして「音域」を導入していたが, より詳細に基本周波数で表すように変更したとみなすこともできる.

5. 評価実験 1

提案手法の有効性を示すため, 評価実験を行う. 実験は 20 回繰り返し, 検定を行う.

5.1 実 験 方 法

まず, 単音データベースとして NTTMSA-P1 を使用した. NTTMSA-P1 は, 6 種類 (piano, violin, flute, trumpet, clarinet, contrabass) の実楽器の単独発音を半音ごとに収録したもので, 各楽器音には, 複数の楽器個体・音の強さ・演奏方法が含まれている. 本実験で使用したデータ (総数: 967 個) の内訳を**表 3** に示す.

表 3 のデータから, ランダムに 2 割の音響信号を選び出し, 特徴量テンプレートを作成する. 残りの 8 割の音響信号を入力データとする (すなわち, どれが学習データに使われて, どれが入力データに使われるかが 20 回の繰り返しで毎回変化し, 学習データと入力データに同じ音響信号が使われることはない). また, 実験は従来の手法 (重み値・類似度計算に木下らの式[3] をそのまま適用したもの) でも行った.

次に, ヤマハ製 MIDI 音源 “MU-2000” から作成した音響データベース (音響信号の総数: 637 個, 内訳は **表 4**) に対しても同様の実験を行った. ただし, 表 2 で示した特徴量の分類は NTTMSA-P1 に基づいて決められたものなので, こちらのデータベース用の特徴量の分類を改めて用意した (**表 5**).

5.2 実 験 結 果

各楽器について, 20 回の実験で得られた認識率:

表 3 NTTMSA-P1 の内訳

楽器の種類	piano(299 個), violin(353 個), flute(238 個), trumpet(77 個). 括弧内はデータ数.
楽器個体数	各楽器に対して 2 種類の楽器個体 (たとえば piano であればヤマハ製とバーゼンドルファー製).
音　域	piano:017-108, violin:055-096, flute:059-098, trumpet:052-084. 数字はすべて MIDI ノートナンバー☆を表す.
強　さ	フォルテ, ノーマル, ピアノ.
奏　法	通常の奏法 (すべての楽器), ビブラート奏法 (violin, flute のみ).

注 1 基本的に上記の音域に対して半音ごとに音高を変えて収録されているが, 抜けている音高が存在する.
注 2 同一条件で複数回収録されているものもある.

表 4 MIDI 音源から作成した音響データベースの内訳

楽器の種類	piano(220 個), violin(143 個), flute(132 個), trumpet(142 個). 括弧内はデータ数.
楽器個体数	各楽器に対して 2 種類の楽器個体.
音　域	piano:045-085, piano 以外:060-085. 数字はすべて MIDI ノートナンバー☆を表す.
強　さ	100, 64, 40. 数字はベロシティを表す.
奏　法	通常の奏法のみ.

注 基本的に上記の音域に対して半音ごとに音高を変えて収録されているが, 抜けている音高が存在する.

表 5 MIDI 音源の場合の各特徴量の分類

第 1 階層

	減衰系楽器	持続系楽器
i)	無相関型	無相関型
ii)	連続変化型	無相関型
iii)	連続変化型	連続変化型

第 2 階層

	弦楽器	管楽器
i)	連続変化型	連続変化型
ii)	連続変化型	無相関型
iii)	連続変化型	無相関型

第 3 階層

主成分	木管楽器	金管楽器
第 1	離散変化型	無相関型
第 2	無相関型	連続変化型
第 3	連続変化型	連続変化型
第 4	無相関型	無相関型
第 5	連続変化型	離散変化型
第 6	連続変化型	無相関型

$$(\text{楽器 } A \text{ の認識率}) = \frac{(\text{正しく同定された } A \text{ の個数})}{(\text{入力データ中の } A \text{ の総数})}$$

の平均値と標準偏差を**表 6** に示す. 提案手法が従来の手法に比べて有効であることを示すため, 検定を行う. 帰無仮説 H_0 とその対立仮説 H_1 をそれぞれ

H_0 : (提案手法の認識率) $\leq$ (従来の手法の認識率)

H_1 : (提案手法の認識率) $>$ (従来の手法の認識率)

とおく. このとき, 帰無仮説 H_0 が棄却されれば対立仮説 H_1 が採択され, 提案手法の有効性が示される.

各楽器に対する検定統計量を表 6 に示す. 有意水準

☆ MIDI ノートナンバーは, MIDI で音高を表すのに使用される情報で, いわゆる中央の C を 60 とし, 半音ごとに 1 ずつ異なる自然数が割り当てられている.

表 6 各楽器の認識率

a. NTTMSA-P1

	提案手法		従来の手法		検定
	平均値	標準偏差	平均値	標準偏差	統計量
piano	.9539	.0378	.9507	.0307	1.2591
violin	.8695	.0313	.7107	.0609	18.0797
flute	.5860	.0808	.5417	.0779	6.8603
trumpet	.6036	.1134	.5499	.1023	8.1049

b. MIDI 音源による音響データベース

	提案手法		従来の手法		検定
	平均値	標準偏差	平均値	標準偏差	統計量
piano	1.0000	0	1.0000	0	NaN
violin	.6377	.0766	.5905	.0828	2.3819
flute	.8588	.0574	.7362	.0747	8.1939
trumpet	.5298	.0592	.2983	.0952	13.0803

を 2.5%とすると, 帰無仮説 H_0 の棄却域は $(2.086, \infty)$ となり, piano 以外のすべての楽器に対して, 帰無仮説 H_0 は棄却される. よって, 提案手法の有効性が示された. piano で提案手法の有効性が現れなかった原因として, 従来の手法で十分に高い認識率を実現していることが考えられる.

5.3 考　　察

本節では, 特徴量の分類の認識率に対する影響や, 階層ごとの認識率について考察する.

5.3.1 特徴量の分類に対する考察

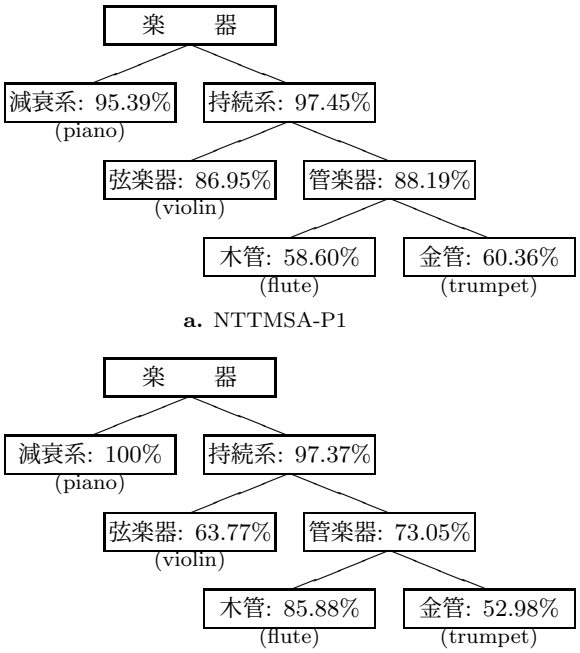
NTTMSA-P1 では, 3 つの階層のうち, 唯一第 2 階層に関しては, 抽出される特徴量に「無相関型特徴量」は 1 つもない. 「無相関型特徴量」は, 音高による音色変化が明確でない特徴量であり, これが少なければ少ないほど提案手法の有効性は大きくなることが予想される. 実際, violin の認識率が最もよく改善されており (約 16%), 検定統計量も際立って大きい (表 6). よって, この予想は実験結果とよく一致している.

また, MIDI 音源用の分類 (表 5) は, NTTMSA-P1 用の分類 (表 2) よりも「無相関型特徴量」が多い. MIDI 音源では, 同じ音響信号を複数の音高に使いまわしており, これが原因と考えられる.

5.3.2 階層ごとの認識率に対する考察

階層ごとの認識率を**図 6** に示す. これより, たとえば piano, flute, violin の 3 重奏に対する音源の弁別のための音源同定手法として用いた場合, NTTMSA-P1 では 86.95%, MIDI 音源では 63.77%の認識率が期待できる☆.

図 6b. では, 「木管楽器」の認識率が「管楽器」の認識率を上回っている. これは, 「金管楽器」の認識率の低さに引っ張られて, 「管楽器」の認識率が低くなっているためである. 「木管楽器」の認識率は, 入力データ中の木管楽器の総数が分母, 管楽器の階層で正しく認識されてさらに木管楽器と正しく同定された個数が分子の割合として求まるのに対し, 「管楽器」の認識率は, 木管楽器と金管楽器の総数が分母, 管楽器と正しく同定された個数が分子の割合となる. そのため, 管楽器と正しく同定されたものの多くが木管楽器のときには, 「木管楽器」の認識率で, 分母の木管楽器の総数に対して分子が大きな割合となるのに対し, 「管楽器」の認識率で, 分母の木管楽器と金管楽器の総数に対して分子は小さな割合となるのである.



a. NTTMSA-P1

b. MIDI 音源による音響データベース

図 6 提案手法による階層ごとの認識率. 各数字は, 該当楽器が該当楽器として同定された割合を表す. **b.** では「木管楽器」の認識率が「管楽器」の認識率を上回っている. これは, 「管楽器」の認識率が, 木管楽器と金管楽器の総数に対する, 同定結果が管楽器である割合であり, 「金管楽器」の認識率の低さが「管楽器」の認識率を引っ張っているからである.

6. 評価実験 2

対象でない楽器 (音源同定システムにとって未知の楽器) を入力したときの動作を評価するため, 実験を行った. 実験方法と実験結果について以下に示す.

6.1 実 験 方 法

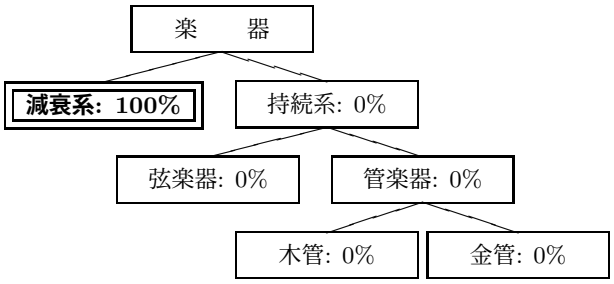
MIDI 音源から作成した音響データベースについてのみ実験を行った. また, 未知の楽器として, classical guitar(減衰系楽器) と clarinet(持続系楽器-管楽器-木管楽器) を使用した.

表 4 で示した音響データベースのすべての音響信号を使って特徴量テンプレートを作成する. 次に, 同音源から classical guitar と clarinet の代表的な MIDI プログラム番号を選び, 入力用の音響データベースを作

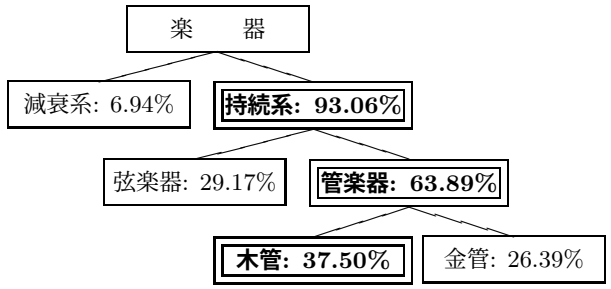
☆ ただし, 本稿の実験では, 混合音は考慮されていないため, 調波構造の推定などの部分で一部改良が必要となる.

表 7 評価実験 2 のための入力用音響データベースの内訳

楽器の種類	classical guitar, clarinet.
楽器個体数	各楽器に対して 1 種類の楽器個体.
音　域	classical guitar:045-085, clarinet:060-085. 数字はすべて MIDI ノートナンバーを表す.
強　さ	100, 64, 40. 数字はベロシティを表す.
奏　法	通常の奏法のみ.



a. classical guitar



b. clarinet

図 7 未知楽器に対する同定結果. この図は図 6 と異なり, 各数字は認識率ではなく, 対象の未知楽器が該当カテゴリーとして同定された割合を表す. そのため, 隣り合う 2 つの数字の和は, それらの親のカテゴリーの数字に等しい. (太字は正解を表す.)

成した. 内訳を**表 7** に示す. 実験は提案手法に対してのみ行い, 各階層に識別される割合を求めた.

6.2 実験結果

実験の結果, 各階層に識別された割合を**図 7** に示す. classical guitar に関しては 100%正しく認識されており, clarinet に関しては第 2 階層までであれば 63.89%の割合で正しく認識された. clarinet が正しく「木管楽器」として認識されたのは 37.50%だが, clarinet は木管楽器ではあるが, リードという発音機構の面で flute と異なるため☆, やむを得ない結果といえる. これは, 今後より多くの楽器に対応するにはカテゴリーの詳細化が必要であることを示唆している.

7. おわりに

本稿では, 同一楽器であっても音高によって音色が変化することに着目し, 音源同定に使われる類似度, および類似度を計算するのに使われる代表値と散らばりの尺度を, 入力信号の基本周波数に関する関数として定義した. その際, 各特徴量を 3 つに分類し, 分類に適した方法で上記の値を計算した. それにより, 認識率の改善に成功した.

また, 階層的な表現を導入することで, 未知の楽器が入力された場合でも一定の階層まで同定が可能であることを確認した. さらに, 本稿の階層的な表現に基づいて楽器音に関するオントロジーを構築することの有用性も指摘した.

今後は, 手動で行った特徴量の 3 つの分類を自動化するだけでなく, より多くの楽器・奏法に対応できるよう, 本手法を拡張していく予定である. また, より認識率を上げるために, 他の特徴量を導入することも検討する. さらに, 打楽器音への拡張, 混合音への適用にも取り組んでいく.

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基盤研究 (B) 第 12480090 号による. また, 音響信号データ NTTMSA-P1 の使用許可を下さった NTT コミュニケーション科学基礎研究所に感謝する. 最後に, 有益な助言を下さった中臺 一博氏に感謝する.

☆ clarinet にはリードがあり, trumpet は両唇がリードの役割をする (リップリードと呼ばれる) のに対して, flute にはそれに相当するものがない (空気流を変化させる機構をエアリードというが, 空気流の変化のさせ方は上 2 つとは異なる)[10].

参 考 文 献

1) 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 (D-II), **J79-D-II**, 11, pp.1751-1761, 1996.
2) 柏野 邦夫, 木下 智義, 中臺 一博, 田中 英彦: “音楽情景分析の処理モデル OPTIMA における和音の認識”, 信学論 (D-II), **J79-D-II**, 11, pp.1762-1770, 1996.
3) 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 (D-II), **J83-D-II**, 4, pp.1073-1081, 2000.
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 (D-II), **J-81-D-II**, 7, pp.1510-1517, 1998.
5) K. D. Martin: “Sound-Source Recognition: A Theory and Computional Model”, PhD Thesis, MIT, 1999.
6) 北原 鉄朗, 後藤 真孝, 奥乃 博: “楽器音オントロジー作成のための楽器音特徴抽出”, 第 62 回情処全大, 4M-5, 2001.
7) 中谷智広, 奥乃 博: “音オントロジーに基づいた音環境理解システムの統合”, 人工知能学会誌, **14**, 6, pp.1072-1079, 1999.
8) 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”, 日本音響学会誌, **33**, 6, pp.291-300, 1977.
9) 早坂 寿雄: “楽器の科学”, 電子情報通信学会, 1992.
10) H. F. Olson 著, 平岡 正徳訳: “音楽工学”, 誠文堂新光社, 1969.
11) 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.
12) 南 茂夫: “科学計測のための波形データ処理”, CQ 出版, 1986.