

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

木下 智義[†] 坂井 修一[†] 田中 英彦[†]

Musical Source Identification Based on Frequency Component Features

Tomoyoshi KINOSHITA[†], Shuichi SAKAI[†], and Hidehiko TANAKA[†]

あらまし 音響信号により外界の事象を理解する聴覚的情景分析に関して、従来多くの研究がなされてきた。特に対象を音楽に絞った場合、自動採譜等の実現を目指し、いくつかの研究例がある。その1つとして、筆者らはこれまでに音楽音響信号を対象とした聴覚的情景分析の処理モデルOPTIMAを提案し、その実験システムを構築した。しかしながら、その認識精度は実用上十分とは言えず、その改善が課題となっている。本論文では、従来の処理の問題点である周波数成分の重なりに対する脆弱性を改善するための新たな処理を提案する。本手法では、周波数成分が重なった時の特徴に合わせて特徴量を分類し、それに応じて重なりのある周波数成分の特徴量を適応的に変化させ、音源同定処理を行う。また、各特徴量の音源同定の際の手掛かりとしての重要度を計算し、同定処理に導入した。評価実験の結果、処理精度の向上が確認され、提案する処理の有効性が明らかになった。

キーワード 聴覚的情景分析、音源同定、音源分離、自動採譜

1. はじめに

聴覚を通じて外界を理解する機能は聴覚情景分析と呼ばれる[1]。自律ロボットなどのように、人間と同等の機能を自動化する際には、聴覚情景分析の計算機上での実現が必要となる。そのため、近年その実現を目指して研究が進められている[2], [3], [5]。

聴覚情景分析のうち、処理対象を音楽に特化した場合を音楽情景分析と呼ぶ。音楽情景分析の機能として、演奏の音響信号からの楽器ごとの音符列、リズム・拍の情報、あるいはより高度な機能として演奏者の意図などの抽出が挙げられるが、この際、音響信号を楽器ごとに分離する処理(音源分離)と、分離された信号の音源名を同定する処理(音源同定)が必須である。しかしながら、特に複数音を対象とした音源同定の場合、それぞれの音が互いに重なりあって干渉することにより、複雑な波形や周波数スペクトルをもつため、高精度の処理は困難であった。

音源同定を試みた例として、筆者らは既に音楽情景分析の処理モデルOPTIMAを提案し[6], [7]、その実験システムを構築した。このシステムでは、複数の異種楽器による演奏のモノラル音響信号を入力として、

楽器ごとの音符列と和音名を出力する。しかしながら、その処理精度は実用上十分であるとは言えず、改善が課題となっていた。

OPTIMAにおける音源同定処理では、各周波数成分の物理的な特徴量を抽出した上で、主成分分析、判別分析といった統計的な処理が用いられている。ここでは複数の単音に由来する周波数成分が重なった場合においても同様の処理を行っている。ところが、周波数成分の重なりが起った場合、それぞれの成分が干渉しあい、そこから得られる特徴量は大きく変動する。この変動により、音源同定に失敗する事例が多くなっている。そこで本論文では、特徴量を3種類に分類した後、周波数成分の重なりが存在した場合に、この分類に従って特徴量の再計算を行い、音源同定における誤りを軽減する新たな処理機構を提案する。

音源同定を扱った他の研究では、最近のものとして、柏野らによる波形レベルでのテンプレートマッチングを用いた手法がある[8]。この手法では、楽器の個体差を吸収するためにテンプレートのフィルタリングや位相トラッキングの処理を行う。これにより楽器個体間の差を吸収しているが、その精度はベンチマークデータに対して約70%となっており、十分な精度とは言えない。

また、三輪らによる、パワーの時間変化のみから音

[†] 東京大学大学院工学系研究科, 東京都
Graduate School of Engineering, The University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan, 113-8656

源の推定を試みた例もある [4] . この手法では , その評価実験において , 同定の対象となる楽器音と比較対象として用いられる基準波形は同一のものをを用いている . また , 周波数成分の重なりが起りにくい音高の組合せとなっている . そのため , 実際の楽器音において想定されるような , 個体間の音色の差や , 周波数成分の重なりによる各成分の変形に対する頑健性には疑問が残る .

本論文では , 2. で複数音を対象とした音源同定における問題点を明らかにしたのち , 3. で新たな手法を提案する . 評価実験については 4. で述べる . 5. で考察を加えた後 , 6. にて結論を述べる .

2. 複数音の演奏に対する音源同定

一般に , ピアノやクラリネットなどの弦楽器 , 管楽器による楽器音は , 楽器の構造から周波数成分が調波構造をもつ傾向にある . また , 複数の音によって協和的なハーモニーを奏でる場合には , それぞれの音の周波数は整数倍の関係になっている . 以上の 2 点により , 調和的な楽器による協和的な演奏では , 複数の単音に由来する周波数成分が同時に同周波数に存在することが多くなる .

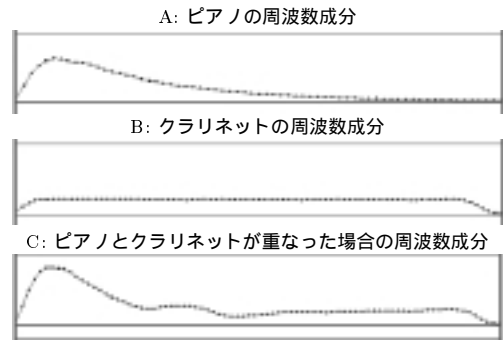
周波数成分が重なりあうと , パワーの小さい成分がパワーの大きい成分に埋もれる形になったり , また干渉によってパワーの時間的変化等の特徴量が変動する . 図 1 はピアノとクラリネットの周波数成分が重なった場合の干渉の様子を示している .

従来の処理では , このような周波数成分の変形を考慮した処理が行われておらず , 音源同定の失敗の原因となっていた . そこで , 本論文ではこの問題を解決するために , 周波数成分が重なった時の挙動により特徴量を 3 種類に分類し , 重なりに応じて入力の特徴量を再計算することにより上述の問題を回避する . この処理の詳細は 3. 3 節にて述べる .

3. 処理の構成

本論文で提案する処理は , 図 2 に示すように , 7 つの処理ブロックと 1 つの知識ベースからなる . 本論文ではこの知識ベースを特徴量テンプレートと呼ぶ .

入力音響信号は最初に前処理 (Preprocess) 部において時間周波数解析され , 次いでその結果から周波数成分が形成される . 単音形成 (Sound Formation) 部では , 前処理部において得られた周波数成分に対してクラスタリングを行う . ここでは , 各クラスタが単音



各音の基音に相当する周波数成分のパワー包絡線形状 . 横軸が時間を , 縦軸がパワーを表す .

- A ピアノの周波数成分のパワーは , 急峻な立上りの後 , 緩やかに減衰する .
- B クラリネットの場合 , 立上りの後は一定のパワー値を保ち , 台形状の包絡線形状となる .
- C 2 つの周波数成分が重なると , 立上り部ではピアノの急峻な立上りが目立ち , 定常部ではクラリネットの平坦な特徴が目立つようになる . また , 立上りから定常部への過渡部では , 干渉により複雑な包絡線形状となっている .

図 1 周波数成分の重なりの影響 .

Fig.1 Influence of overlapped frequency components.

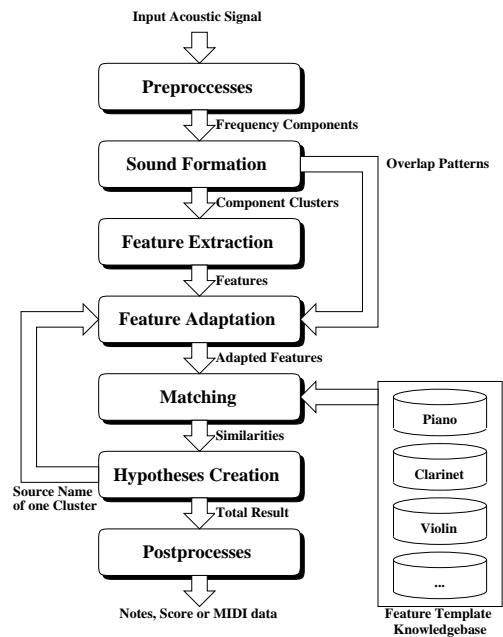


図 2 処理の構成図 .

Fig.2 Block diagram of proposed process.

に相当する . また , 単音形成と同時に周波数成分の重なりパターンが抽出される . 続く特徴抽出 (Feature

表 1 特徴量テンプレートの例 .
Table 1 Example of feature template database.

音源名	立上りの強さ	第 1 成分のパワー値	中心周波数	...
ピアノ	0.317	0.487	2.571	...
ピアノ	0.539	0.555	2.242	...
...
クラリネット	0.258	0.634	2.879	...
...

Extraction) 部では、それぞれの周波数成分から特徴量が得られる。この特徴量は、単音形成部で得られた重なりパターンに従い、特徴量適応 (Feature Adaptation) 部で変形される。マッチング (Matching) 部では、変形された特徴量と特徴量テンプレートに格納されている特徴量の間の比較を行い、類似度が計算される。この類似度は仮説生成 (Hypotheses Creation) 部へと送られ、単音仮説が形成される。全てのクラスタについての音源同定が完了するまで特徴量適応部から仮説生成部の間の処理は繰り返し行われることになる。最後に後処理 (Postprocesses) 部にて単音データや楽譜が作成される。各処理部の動作に関する詳細は後述する。

特徴量テンプレートは、それぞれが 1 つの単音に相当するレコードからなる。各レコードには音源名と、特徴量の値のリストが含まれる。表 1 にその例を示す。

3.1 前処理および単音形成

前処理部においては、最初に入力信号に対して時間周波数解析を行い、続いてここから周波数成分を抽出する。これらの処理においては、IIR フィルタバンクと狭平面回帰法を用いた [6]。また、特徴量テンプレートを作成する際にも、モデルとなる単音データに対して同様の処理を施す。

続いて、単音形成クラスタリングを行う。この処理において、周波数成分は単音ごとにクラスタリングされる。ここでは柏野の手法 [6] を用いた。この手法では、調和性や複数の周波数成分の間の立上り時刻のずれを抽出し、この結果に応じて同一の単音によると思われる周波数成分どうしを集めてクラスタリングを行う。

本論文ではこれに加えて周波数成分の重なりパターンを抽出した。ここで、重なりパターンとは、複数のクラスタに属する周波数成分と、この成分が属するクラスタの組の集合として定義される (図 3)。

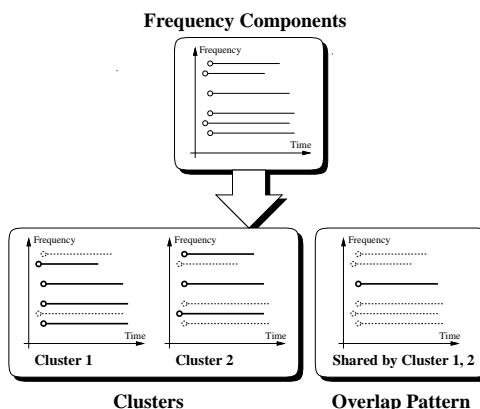


図 3 単音形成クラスタリングと重なりパターンの抽出 .
Fig.3 Sound formation and extraction of overlap patterns.

表 2 周波数成分の特徴量 .

Table 2 Feature of frequency components.

- 各周波数成分のパワー値 (各パワー値の、全成分のパワー値の総和に対する比)
- 各周波数成分の立上りの強さ (隣接するサンプル点のパワー値の差の最大値を、成分の最大パワー値で割ることで正規化したもの)
- 奇数次倍音と偶数次倍音のパワー総計の比
- 中心周波数 (各成分の周波数値のパワーを重みとした加重平均を、基音の周波数で割ることで正規化したもの)
- パワー値の時間方向対称性 (各周波数成分のパワー包絡線について離散正弦変換を施した場合の、偶数次と奇数次の係数の総和の比)
- パワー値の時間方向振動性 (各周波数成分のパワー包絡線について離散正弦変換を施した場合の、低次 (6 次まで) と高次 (7 次以上) の係数の総和の比)
- 各周波数成分のパワー包絡線をパワー値の分布とみた時の、歪度と尖度

3.2 特徴量抽出

次に、前段階で得られた周波数成分から特徴量を抽出する。本論文では特徴量はパワー包絡線の形状や、立上りの強さ、あるいは各周波数成分のパワーの比といった物理的な量を用いている。表 2 にその一覧を示す。

一部の高調波について、そのパワー値が小さい場合には、この周波数成分から得られた特徴量には意味がないものとして無効化し、後述するマッチング部において類似度計算の対象から外すものとする。

3.3 特徴量の適応処理

実音楽の場合、通常は同時に複数の単音が存在し、

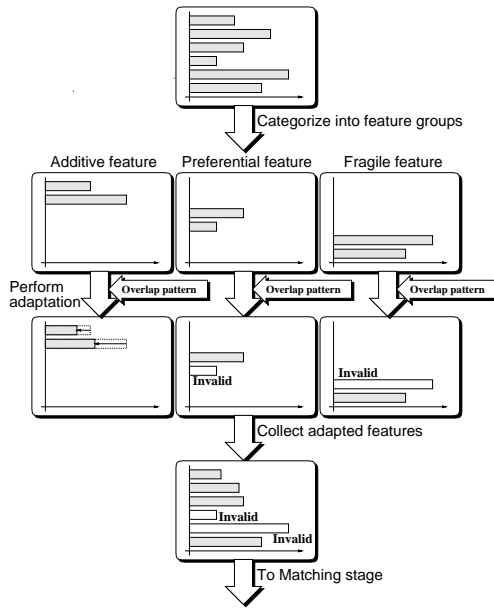


図4 周波数成分特徴量の適応機構。
Fig. 4 Adaptation mechanism of features.

また、周波数成分のうちいくつかは重なりあうことが多い。これにより、それぞれの周波数成分は変形され、特徴量は変化してしまう。そのため、重なりのある周波数成分については、その特徴量をそのまま用いて音源同定すると誤認識の原因となる。

そこで本論文では、周波数成分の重なりパターンと、各周波数成分の性質により特徴量の変形を行うものとする。ここでは図4に沿ってこの流れを検討する。

3.3.1 特徴量の分類

まず、事前に特徴量をその特質によって3種類に分類した。以下にその分類を示す。

(1) 加算特徴量 (Additive feature)

周波数成分が重なった時に、その周波数成分から計算される特徴量も概ねそれぞれ単独の場合の和になるもの(例: 周波数成分のパワー値)

(2) 優先特徴量 (Preferential feature)

重なった周波数成分の特徴量のうち、最大もしくは最小の値が全体の特徴量として得られるもの(例: 立上りの強さ)

(3) 崩壊特徴量 (Fragile feature)

周波数成分が重なった場合、得られた特徴量が意味をなさなくなるもの(例: パワー包絡線形状の対称性)

3.3.2 適応処理

周波数成分がただ1つのクラスに属する時には前段で計算された特徴量がそのまま用いられる。一方、複数の周波数成分が重なった場合には、前項による分類に従って、特徴量の再計算を行う。再計算は以下のように行われる。

(1) 加算特徴量

適応処理は以下のアルゴリズムに従って行われる。

If 周波数成分が属するクラスのうち1つについて、既に音源名が決定されている。

Then

既に決定されている音源の特徴量テンプレートから特徴量を得、入力信号から計算された特徴量から引く。

Else

特徴量の再計算は行われない。

(2) 優先特徴量

以下に述べるようなアルゴリズムに従って適応処理をする。

If 周波数成分が属するクラスのうち1つについて、既に音源名が決定されている

Then

既に決定されている音源の特徴量テンプレートから特徴量を得る。さらに、その値に応じて、以下のように処理を進める。

If 入力信号から計算された特徴量と、テンプレートから得られた特徴量が十分近い値となっている。

Then

入力からの特徴量は、既に決定している音源によるものと判断し、特徴量を無効にしてマッチング部での類似度計算の対象から外す。

Else

特徴量の再計算は行われない。

Else

特徴量の再計算は行われない。

(3) 崩壊特徴量

既に特徴量が意味をなさないと判断されると判断されるため、特徴量を無効にし、マッチング部での類似度計算の対象から外す。

3.4 マッチング

音源同定は、適応処理を施した特徴量とテンプレート中の特徴量との類似度を計算することで行う。

3.4.1 重み値の計算

周波数成分の特徴量が、音源同定の手掛かりとなることは示されている [6] もの、全ての特徴量に、手掛かりとして同程度の意味があるとは限らない。実際、クラリネットは、偶数次高調波のパワーが非常に小さいという特徴があり、この点はクラリネットの同定において他の特徴量と比べて大きな手掛かりとなる。そこで、本論文では事前に各音源ごとに、特徴量の重要度を計算した。

まず、事前に特徴量テンプレートに格納されている各特徴量について、音源ごとに平均と標準偏差を計算する。続いて、以下の式に従って各特徴量の重み値を計算する。本論文では、上付文字で特徴量の種類を、下付文字で音源の種類を表すものとする。

$$w_{s,t}^i = P \left(|X| \leq \frac{|\mu_t^i - \mu_s^i|}{\sigma_s^i} \right) \quad (1)$$

$$W_s^i = \sqrt{\frac{1}{|S|-1} \sum_{t \in S, t \neq s} w_{s,t}^i{}^2} \quad (2)$$

ここで、 S は音源名の集合を表す ($= \{Piano, Clarinet, \dots\}$)。また、 s は個々の音源を表すものとする ($s \in S$)。 μ, σ はそれぞれ特徴量の平均値と標準偏差で、 P は、次式のような、標準正規分布における確率値を表す。

$$P(|X| \leq z) = \int_{-z}^z (1/\sqrt{2}) \exp(-x^2/2) dx \quad (3)$$

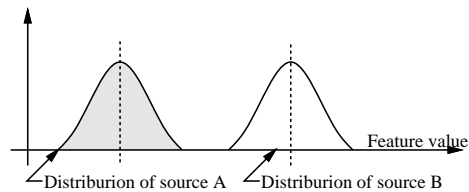
今、 D_s^i を音源 s の i 番目の特徴量の分布とする。上記の式により、 W_s^i は D_s^i が他の音源の分布と離れている場合に大きな値となり、逆に近い時に小さな値となる。例えば、 D_s^i が他の音源の D_s^i と十分離れていれば W_s^i は 1 となり、全ての音源について D_s^i が同じ位置に分布する時には W_s^i は 0 となる。こうして、大きな W_s^i を持つ特徴量 i は音源同定の手掛かりとして重視される (図 5)。こうして得られた W_s^i は次段の類似度計算にて用いられる。

3.4.2 類似度計算

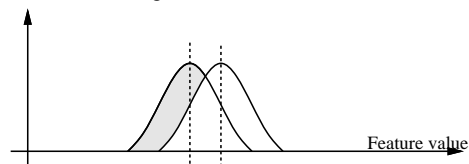
次いで、入力信号から得られた特徴量とテンプレートから得られる特徴量間の類似度を計算する。この類似度が音源同定の根拠として用いられる。

まず、類似度そのものの計算に先立ち、各特徴量ごとに入力信号の特徴量とテンプレートの特徴量間の距離を計算する。距離は以下のように計算される。

Case 1: Large weight



Case 2: Small weight



Case 1: 2つの特徴量の分布が十分に離れている場合、この特徴は音源同定の手掛かりとして重要であると判断し、大きな重み値を与える。

Case 2: 分布が近い場合には、音源同定の手掛かりにはならないと判断する。

図 5 特徴量の分布に応じた重み値の計算。

Fig. 5 Weighting based on the distribution of feature values.

$$d_s^i = P \left(|X| \geq \frac{|f^i - \mu_s^i|}{\sigma_s^i} \right) \quad (4)$$

ここで、 d_s^i, f^i はそれぞれ i 番目の特徴量に関する入力とテンプレート間の距離、入力信号から得られた i 番目の特徴量の値を意味する。 P は、前項と同様、標準正規確率分布における確率値である。

最後に、類似度を以下の式に従って計算する。

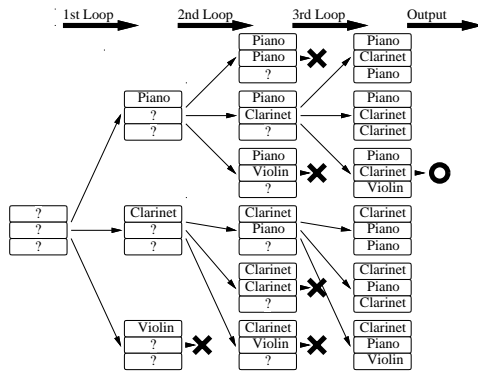
$$R_s = \exp \left(\frac{\sum_i W_s^i \log d_s^i}{\sum_i W_s^i} \right) \quad (5)$$

ここで、 W_s^i と d_s^i はこれまでの処理で得られた値である。

この式において、 i は各特徴量を表すが、特徴量抽出や適応処理の段階でこの特徴量が無効であると判断されている場合には、 \sum の計算から除外される。

3.5 仮説生成

マッチング処理の後、最も低い基本周波数をもつクラスタについて、その音源名を確定させる。この一部の音源名が確定したデータは適応処理部へフィードバックされる。フィードバック後の適応処理部では、確定した音源の情報を用いて再度適応処理を行うこととなる。実際には音源名を一意には確定させず、複数



各矩形がクラスに相当する。処理がループすることにより、音源名を一つずつ与えていくが、途中、類似度の低いものは候補から除外される。

図6 処理のフィードバックの例
Fig.6 Example of process feedback.

の候補を作成してそれぞれについてフィードバックを行うことになる。

最初に行われたマッチング処理で誤りが生じた場合でも、フィードバック後の処理では前段で確定した音源名を用いて処理を行うため、低い類似度が与えられることが期待できる。

例えば、低いピアノの単音と高いクラリネットの単音からなる信号を入力した場合を考える。最初の処理において、周波数成分の重なりのために、ピアノの単音に対してピアノよりもクラリネットとの類似度の方が大きな値となった場合でも、フィードバック後の処理で「クラリネット+クラリネット」よりも、「ピアノ+クラリネット」の方に高い類似度が付与されれば、誤りを修正することができる。実際、特徴量変形部に低い音がピアノであるとして処理を進めた場合には、高い音に含まれるピアノの特徴を除外することができるため、高い音に対するクラリネットとの類似度が大きな値になることが予想される。

途中、候補の組合せが一定数を越えた場合には、類似度の低い候補を除外することで、組合せの爆発を防ぐ。

最終的に、類似度の値が大きくなった候補を出力とし、単音仮説を生成する。各単音仮説は複数の単音を含み、それぞれの単音は開始時刻、継続時間、音高、音源名の情報をもつ(図6)。

3.6 後処理

OPTIMAの枠組において、仮説生成部で得られた単音仮説は他の処理モジュールから得られる確率情報

表3 各音源における重み値の大きい特徴量上位3項
Table 3 Top three large weighted features.

音源名	特徴量	W_s^i
ピアノ	基本波のパワー値の時間方向対称性	0.956
	2倍音のパワー値の時間方向対称性	0.943
	基本波のパワー値分布の歪度	0.937
クラリネット	基本波のパワー値分布の尖度	0.721
	3倍音のパワー値の時間方向対称性	0.719
	2倍音のパワー値	0.703
ヴァイオリン	基本波の立上りの強さ	0.899
	2倍音の立上りの強さ	0.869
	基本波のパワー値分布の尖度	0.771

と統合される[6]。この統合処理により出力された単音仮説に含まれる誤りが訂正されることが期待される。

4. 評価実験

本論文で提案した処理を検証するために、2種類の評価実験を行った。まず、特徴量の重み値の計算を行い、その結果の妥当性を確認した。続いてベンチマークデータに対する音源同定処理の精度を計算し、適応処理の有無による差を評価した。

4.1 重み値計算の評価

まず、表3にマッチング部における W_s^i の計算結果を示す。

この結果は、人間の通常の直感に合ったものとなっている。ピアノの周波数成分は鋭く立上り、緩やかに減衰する。またそのパワーは立上り付近に主に分布し、減衰域でのパワー値は小さなものとなっている。上表での時間方向対称性と歪度はいずれもピアノのパワー包絡線形状の非対称性を反映したものとなっている。クラリネットは、偶数次の高調波のパワーが非常に小さく、またそれぞれの成分は台形状のパワー包絡線をもつ。従って、パワー分布の尖度や2倍音のパワー値の小ささがその特徴として現れている。ヴァイオリンの立上りは本論文で用いた音源の中では比較的緩やかであった。それを反映して、立上りの強さの重み値が大きなものとして得られている(図7)。

4.2 音源同定処理の評価

ここでは、ベンチマークデータとして用意したランダムノートパターンを対象に、音源同定処理の認識精度の評価を行った。

ランダムノートパターンは、同時に立上る3つの単音の組の集合である。各単音の音高と音源名はランダムに決定されている。また、周波数成分の重なりに応じ以下のように分類した。

クラス1: 1つの単音の2倍音の成分が、別の単音

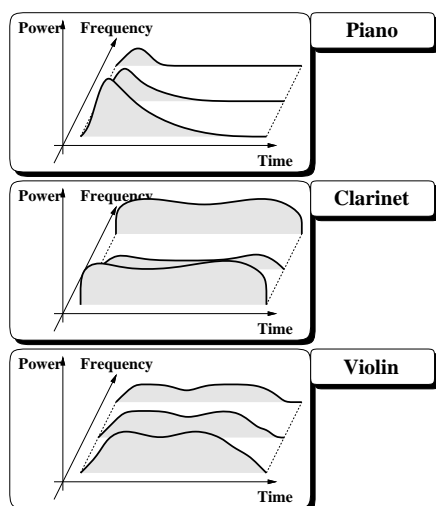


図 7 各音源の周波数成分の標準的な形 .
Fig.7 Typical shapes of sound source's frequency component.

の基本波と重なるような単音の組 . この場合 , 1 つの単音の周波数成分の全てが別の単音の周波数成分と重なることになる . 本論文で分類した 3 つのクラスの中では最も認識が困難になる .

クラス 2: 1 つの単音の 3 倍音の成分が , 別の単音の 2 倍音成分と重なるような単音の組 . これら 2 つの単音は基本周波数の比が 2 : 3 で , 完全 5 度の関係になっている .

クラス 3: クラス 1 にもクラス 2 にも分類されないもの . この場合 , 周波数成分の重なりは比較的少ない .

本実験では , 各クラス 300 パターンずつ用意し , 合計 900 パターンに対して 3 音の加算合成を行ったものを入力として処理を行った . 入力信号とテンプレートを作成するデータとして音響信号データ NTTMSA-P1 の単音データを用いた . この信号データは , 単一の実楽器による演奏を録音したものである . また , 処理対象となる音響信号と , 特徴量テンプレートの作成に用いた音響信号は , 同一の音源名の場合でも別の楽器個体からのものを用いた .

特徴量変形部での優先特徴量の変形においては , 入力の特徴量を音源テンプレートに格納されている特徴量の分布に照らし合わせ , その平均との差が標準偏差の範囲内にある場合に近い値であると判断した .

本論文では入力として用意した単音と , 出力された単音の音高と音源名がともに一致した場合を正解とみ

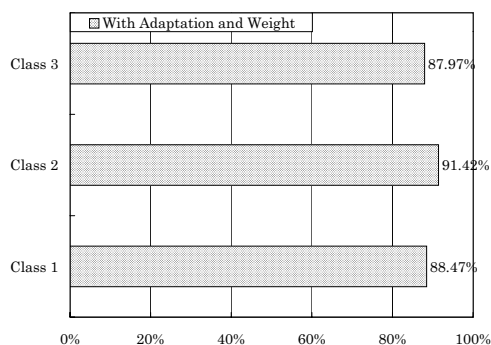


図 8 単音形成のみの認識精度 .
Fig.8 Result of note creation for random note patterns.

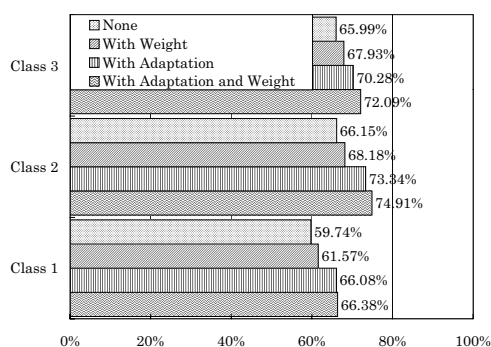


図 9 音源同定も含めた認識精度 .
Fig.9 Result of whole process for random note patterns.

なし , また , 認識精度として再現率と適合率の平均を用いた . すなわち ,

$$\left(\frac{\text{出力中の正解数}}{\text{出力された単音数}} + \frac{\text{出力中の正解数}}{\text{正解中の単音数}} \right) \times \frac{1}{2}$$

である . 図 8-10 に認識結果の精度を示す . ここで , 認識精度の 95% 信頼区間は 1.5% 程度である .

図 8 は , 単音形成精度をクラスごとに表したものである . 単音形成精度は , 音高のみを用いて正誤判定を行ったもので , 本論文における処理では音源同定処理の前に単音形成を行うため , 音源同定処理の精度がこの数値を上回ることはない .

続いて図 9 は , 音源同定精度を表す . ここでは , 音高と音源名の両方が正しかった場合を正解として扱った . 処理の効果を見るために , 重み値処理と特徴量変形を行った場合と行わなかった場合の 4 つの場合において精度を求めた . それぞれのグラフの意味は表 4 の

表4 ランダムノート実験におけるパラメータ.
Table 4 Parameters in random-note experiment.

	重み値	特徴量変形
None	×	×
With Weight		×
With Adaptation	×	
With Adaptation and Weight		

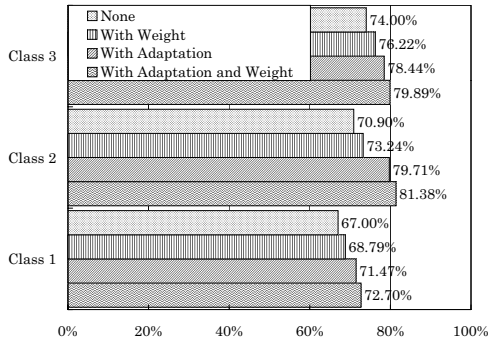


図10 音源同定処理のみの認識精度.

Fig. 10 Result of source identification when the pitch is given beforehand.

通りである.

図9に示した数値は、音源同定処理のみならず、単音形成処理における誤りも含んだものとなっている。そこで、あらかじめ単音の開始時刻と音高を与えることにより、音源同定処理のみの精度を計測した。その結果を図10に示す。ここで、4つのグラフの意味は、図9のものと同様である。

5. 検討・考察

本論文では、音源同定処理に、周波数成分の重なりに応じた特徴量の適応処理を行い、また類似度計算における特徴量ごとの重み値を導入することで、精度を向上させることに成功した。その効果は最大で70.90%から81.38%へと、10%程度となった。

一方で、精度自体は改善の余地を多く残している。これは、本論文では適応処理のうち、一部で「何もしない」あるいは「特徴量を無効にする」といった簡単な処理を用いたことが一因として挙げられる。また、演奏に用いたデータと、テンプレートとして用意したデータは、別の楽器個体による演奏のものを用いた。そのため、同一の楽器であっても特徴量に若干の差が存在し、そのために誤認識を生じる例があった。

ところで、本論文で示した実験において、クラス2の結果がクラス3の結果よりも優れる傾向となった。

この原因として以下の点が考えられる。クラス2の演奏データは、クラス3の演奏データに比べて、周波数成分の重なりが生じる傾向にある。逆に周波数成分が重ならないクラス3の演奏データには数多くの周波数成分が存在することとなる。この周波数成分の中には、周波数が近接しているものも含まれ、そのため周波数成分の抽出に失敗した可能性がある。

6. おわりに

本論文では、音源同定処理に特徴量の変形処理と、特徴量ごとの重み値を導入することにより、精度の向上をみた。

図10におけるクラス2の実験は、文献[8]におけるベンチマーク実験と同一のデータを用い、ほぼ同じ条件で行われたものである。この結果に比べても10%程度優れた結果が得られた。

一方で、精度自体は改善の余地を多く残しており、今後は同定の対象となる楽器の種類等をも考慮にいれて、楽器ごとに特有の適応処理を加えるなどの改善を進める必要がある。

また、クラス3の実験結果から、周波数成分抽出部における課題が顕在化した。今後はこの点も考慮して検討を進める予定である。

謝辞

本論文は、文部省科学研究費補助金(課題番号09-07628)による研究成果の一部である。また、音響信号データ NTTMSA-P1の使用許可をいただいた NTTコミュニケーション科学基礎研究所に感謝する。

文献

- [1] A. S. Bregman. Auditory scene analysis. MIT Press, 1990.
- [2] Daniel P. W. Ellis. Prediction-driven computational auditory scene analysis. PhD thesis, Department of Electrical Engineering and Computer Science, M.I.T., 1996.
- [3] Victor R. Lesser, S. Hamid Nawab, and Frank I. Klassner. IPUS: An architecture for the integrated processing and understanding of signals. *Artificial Intelligence*, Vol. 77, pp. 129-171, 1995.
- [4] 三輪多恵子, 田所嘉昭, 斎藤努, “くし形フィルタを利用した採譜のための異楽器音中のピッチ推定,” 信学論(D-II), Vol. J81-DII, No. 9, pp. 1965-1974, 1998.
- [5] 中谷智広, 後藤真孝, 川端豪, 奥乃博, “残差駆動型アーキテクチャの提案と音響ストリーム分離への応用,” 人工知能誌, Vol. 12, No. 1, pp. 111-119, 1997.
- [6] 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), Vol. J79-DII, No. 11, pp. 1751-1761, 1996.

- [7] 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), Vol. J79-DII, No. 11, pp. 1762-1770, 1996.
- [8] 柏野邦夫, 村瀬洋, “適応型混合テンプレートをを用いた音源同定 — 音楽演奏への応用 —,” 信学論(D-II), Vol. J81-DII, No. 7, pp. 1510-1517, 1998.

(平成 xx 年 xx 月 xx 日受付)

木下 智義 (学生員)

平 7 東大・工・電子情報卒。現在同大学院情報工学専攻博士課程在学中。聴覚的情景分析の研究に従事。情報処理学会, 人工知能学会, 日本音響学会, 各会員。日本学術振興会特別研究員。

坂井 修一 (正員)

1958 年生。1981 年東京大学理学部情報科学科卒業。1986 年同大学院情報工学専門課程修了。工学博士。同年, 電子技術総合研究所入所。1991 年 4 月より 1 年間米国 MIT 招聘研究員。1993 年 3 月より 1996 年 2 月まで RWC 超並列アーキテクチャ研究室室長。1996 年 10 月より 1998 年 3 月まで筑波大学助教授(電子・情報工学系)。1998 年 4 月より東京大学助教授(工学系研究科)。計算機システム一般, 特にアーキテクチャ, 並列処理, スケジューリング問題, マルチメディアなどの研究に従事。情報処理学会論文賞(1990 年度), 日本 IBM 科学賞(1991 年), 市村学術賞(1995 年), ICCD Outstanding Paper Award(1995 年)など受賞。情報処理学会, IEEE, ACM, 各会員。

田中 英彦 (正員)

1943 年生。1965 年東京大学工学部電子工学科卒業。1970 年同大学院博士課程修了。工学博士。同年東京大学工学部講師。1971 年助教授。1978 年～1979 年ニューヨーク市立大学客員教授, 1987 年教授, 現在に至る。計算機アーキテクチャ, 並列処理, 人工知能, 自然言語処理, 分散処理, CAD 等に興味を持っている。「非ノイマンコンピュータ」「情報通信システム」著。「計算機アーキテクチャ」「VLSI コンピュータ I, II」「ソフトウェア指向アーキテクチャ」共著。New Generation Computing 編集長。情報処理学会, 人工知能学会, 日本ソフトウェア科学会, IEEE, ACM, 各会員。