2004－MUS－56　(9)
2004／8／3

混合音テンプレートを用いた多重奏の音源同定

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
‡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，混合音から得られる特徴量テンプレート（楽器名ラベル付きの特徴ベクトルの集合）を用いて，多重奏の音源同定を行う．多重奏の音源同定はこれまでにいくつかの研究がなされてきたが，いずれも単一楽器音のみからテンプレートを得ていた．本稿では，特徴量テンプレートを混合音から作成することで，音の混合による特徴変動への対処を試みる．しかし，混合音の組み合わせは無数にあるため，すべてを網羅したテンプレートを作成することは不可能である．本研究では，実際の楽曲の楽譜から混合音を作成することで，現実の楽曲で用いられる混合音を重点的に収集する．三重奏の音響信号を用いた実験の結果，提案手法により音源同定の正解率が 72.3%から 80.8%に改善されることを確認した．

Musical Instrument Identification in Polyphonic Music using a Mixed-sound Template

Tetsuro Kitahara† Masataka Goto‡ Hiroshi G. Okuno†

†Dept. of Intelligence Science and Technology,
Graduate School of Infomatics, Kyoto University
‡National Institute of Advanced Industrial Science and Technology

Abstract This paper describes musical instrument identification in polyphonic music using a feature vector template (i.e., a set of feature vectors with musical instrument labels) obtained from mixtures of sounds. Although there are some studies dealing with polyphonic music, all of them use a template obtained from only solo sounds. In this paper, we construct a template from mixtures of sounds. Because there are infinite combinations of instruments, however, it is impossible to construct the template that covers all of them. To solve this problem, by making mixtures of sounds from a score of an actual musical piece, we obtain only musical instrument combinations appeared in an actual musical piece. Experimental results using trio music show that the proposed method has improved the recognition rate from 72.3% to 80.8%.

1. はじめに

近年，ディジタル音楽配信の普及にともない，膨大な音楽音響信号がインターネット上に蓄積されつつある．膨大な音楽音響信号から効率よく目的の楽曲を検索するには，MPEG-7 などの共通フォーマット下で，検索に効果的な楽曲の内容を記述することが肝要である．なかでも,「どの楽器で演奏しているか」という情報は,「弦楽四重奏」「ピアノソナタ」などのような分類名が用いられることからも分かるように，音楽検索において重要な役割を果たすと考えられる．

音楽音響信号から演奏されている楽器の名前を同定する処理は音源同定と呼ばれ，これまでにさまざまな研究がなされてきた[1)~14)]．これらの研究は，単一音のみを扱うもの[1)~10)]と混合音を扱うもの[11)~14)]とに大きく分けることができる．単一音を扱う研究のほとんどは，単一楽器の孤立発音 (isolated notes) を扱ってきた[1)~6),10)]．これらの研究の特長は，対象とする楽器数が多いことで，10〜30 種類程度の楽器を対象に 70〜80%程度の認識率を実現しているものが多い．文献 7)~10) では単旋律の音源同定も扱っているが，対象とする楽器数は孤立発音よりも少ない．一方，混合音を

扱う研究[11)~14)]では，3〜5 種類程度の楽器を対象に，二〜三重奏の音源同定に取り組んでいるものがほとんどである．文献 11) では，5 種類の楽器を対象に三重奏の楽曲を扱っているが，音源同定部に対する性能評価は行っていない．文献 12)，13) では，3 種類の楽器を対象に三重奏の音源同定を行い，約 70〜80%の認識率を得ている．また，文献 14) では，5 種類の楽器を対象として二重奏の音源同定を行っており，49%の認識率が報告されている．

本研究では，多重奏の音源同定の第 1 段階として，4 種類の楽器を対象に三重奏の音源同定に取り組む．これは，これまでの混合音の音源同定研究のなかでは標準的な問題設定であり，現状の技術レベルで取り組むには適していると考えられる．また，この程度の複雑さの楽曲を扱う場合には，音高の遷移などのトップダウン情報を用いることで音源同定の性能を改善できることが報告されている[15)]が，本稿では，音源同定の基礎技術確立のため，このようなトップダウン情報は用いずに，ボトムアップな方法のみで音源同定を行う．

多重奏（混合音）の音源同定が難しいのは，周波数成分が重複することにより，特徴量が大きく変動するからである．この問題に対して，これまでの研究[12)~14)]ではさまざまな対策がとられてきたが,「単一楽器音のテンプレート（学習データ）を用いて混合音を認識する」という枠組みはどの研究においても共通であった．

本研究では，この問題に対して，混合音から作成したテンプレートを用いる．これにより，学習時と認識時とで同様に周波数成分の重複による特徴変動が起こるため，単一楽器音のテンプレートのみを用いるのに比べてロバストになると期待される．しかし，混合音の組み合わせは非常に多く，すべての組み合わせを網羅的に収集するのは現実的には不可能である．そこで，実際の楽曲の楽譜から混合音を作成することにより，現実の楽曲で用いられる混合音の組み合わせを重点的に収集することを検討する．

以下，**2.** では混合音の音源同定における課題を議論したのち，本研究での解決策を述べる．**3.** では「混合音テンプレート」の詳細について述べ，**4.** で我々の音源同定システムの処理の流れを述べる．**5.** で評価実験について述べ，最後に **6.** でまとめをする．

2. 混合音に対する音源同定

多重奏（混合音）の音源同定が難しい主たる原因は，周波数成分が重複することにより，そこに含まれる各単音☆の音響的特徴を正確に抽出できないことにある．もしも，音源分離技術により，混合音から各単音を完全に分離できるのであれば，混合音に対する音源同定は，単一音の音源同定の問題に帰着することができる．しかし，実際には，周波数成分の重複が頻繁に発生するため，混合音を歪みなく分離するのは不可能である．そのため，単一音を対象とした音源同定手法では，周波数成分の重複に伴う特徴変動によって，性能が大きく低下する．

周波数成分の重複に伴う特徴変動によって，音源同定の性能が大きく低下するのは，特徴量がどう変動するのかを音源同定システムが知らないからである．すなわち，周波数成分の重複によって各特徴量はどのように変動するかを，システムが十分に学習していれば，混合音に対しても高い性能を示すと考えられる．それには，特徴量がどのように変動するかをシステムが知る仕組みが必要である．また，よりロバストな音源同定を実現するには，変動が大きな特徴量には小さな重みを，変動が小さな特徴量には大きな重みを与える仕組みも必要となる．以上より，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　周波数成分の重複によって変動した特徴量をどのように学習するか．

課題 2　各特徴量に対して，どのように周波数成分の重複による変動の程度に応じた重みを設定するか．

これまで，これらの課題に対していくつかの解決策が提案されてきた．以下に，その代表的なものを示す：

- 適応型混合テンプレート法[12)]では，単音の特徴抽出を行わず，各音源の波形テンプレートの和と入力波形との自乗誤差が最小になるように波形テンプレートを変形させた上で，波形レベルで両者のマッチングを行う．この手法では，単音の特徴抽出を行わないことで，特徴変動の問題を本質的に回避している．しかし，波形レベルでテンプレートマッチングをするには，テンプレートが入力波形に十分に類似している必要がある．その問題を解決するために，FIR フィルタと位相トラッキングによるテンプレート適応処理が提案されているが，これらだけでは実楽器音の多様性を吸収するには限界がある．
- 木下らの周波数成分重なり適応処理[13)]では，変動

☆ 本稿では,「単音」と「単一音」を異なる意味で用いる．前者は，処理の一単位となる音で，楽譜上の一音符に相当する．通常，1 つの単音は調波構造を 1 つだけ持つ．一方，後者は，単音が同時に 1 つしか鳴っていない音を指す．

の仕方で特徴量を 3 つに分類し，周波数成分の重なりがあったときに，この分類にしたがって特徴量の再計算を行う．これは**課題 1** に対応したものであるが，特徴量の分類は手動で行う必要があることや，特徴量の再計算には，同時に鳴っている他の音源の名前がすでに確定している必要があることから，この手法の有効範囲は大きくないと考えられる．また，**課題 2** に対しては，特徴量の重要度を計算する処理を導入しているが，単一音に基づいて計算されており，特徴変動の程度に応じた重み設定という観点には至っていない．

- Eggink ら[14]は，周波数成分重複による特徴変動の問題を，Missing feature theory を用いて解決している．Missing feature theory は，信頼できない特徴量をマスクすることで，事後確率の計算でこのような特徴量を使わないようにする方法である．これは，**課題 1** を扱わずに，**課題 2** において，変動の可能性のある特徴量に重み 0 を与えることで，特徴変動に対処するものである．しかし，Missing feature theory における最大の課題である，マスクすべき特徴量を自動推定する技術は発展途上であり，現状では，この自動推定を行うには，ローカルスペクトルなどの限られた種類の特徴量しか利用できない．

本研究では，これら 2 つの課題を次のように解決する．

解決策 1　混合音からのテンプレート作成

音源同定に用いる特徴量テンプレートを混合音から作成する．特徴量テンプレートとは，楽器名がラベルづけられた特徴ベクトルの集合で，各楽器の特徴空間上の分布の確率密度関数を推定するのに用いられる．このように混合音から特徴量テンプレートを作成することで，学習時と認識時とで同様の特徴変動が起こることになり，性能向上が期待できる．本稿では，混合音から作成した特徴量テンプレートを**混合音テンプレート**と呼ぶ．

解決策 2　クラス内分散・クラス間分散比最大化基準に基づく次元圧縮

混合音から抽出した特徴ベクトルを用いて特徴空間上の分布を形成した場合，周波数成分の重複によって特徴変動が起きると，その特徴量のクラス内分散が大きくなる．そのため，クラス内分散・クラス間分散比（クラス間の分離度に相当）が低下する現象が見られる．そこで，クラス内分散・クラス間分散比最大化基準に基づく次元圧縮法である**線形判別分析**を用いることで，特徴変動の大きな特徴量の重みを小さくする次元圧縮を実現する．

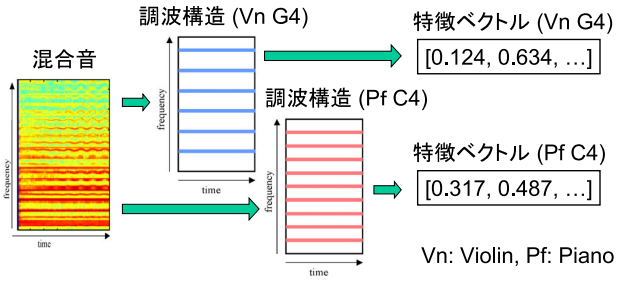

図 1　混合音テンプレート作成の流れ．テンプレート作成に用いる混合音には，含まれている各単音の音高・発音時刻・楽器名がラベルづけされており，これに基づいて各単音の調波構造を抽出する．そして，特徴抽出を行い，得られた特徴ベクトルをテンプレートとして蓄積する．

3.　混合音テンプレート

本研究では，音源同定に用いる特徴量テンプレートを混合音から作成する．テンプレート作成の手順を**図 1** に示す．テンプレート作成の元となる混合音には，含まれる各楽器音の楽器名・音高・発音時刻がラベルづけられている．そのラベルに基づき，各楽器音の調波構造を抽出する．次に，その調波構造に対して特徴抽出を行う．

しかしながら，混合音の組み合わせは非常に多いため，すべての組み合わせを網羅的に収集するのは現実的には不可能である☆．そのため，組合わせ爆発を避ける何らかの方法が必要となる．

本研究では，実際の楽曲の楽譜から混合音を作成することで，現実の楽曲で出現され得る混合音の組み合わせのみを重点的に収集する．混合音の全組み合わせのうち，多くのものは不協和音を生むため，実際の音楽ではほとんど用いられない．そのため，実際の楽曲で用いられる混合音の組み合わせには偏りがあり，ある楽曲で出現しなかった組み合わせは，他の楽曲でも出現しにくいと予想される．そのため，少数の楽曲の楽譜から混合音を作成することで，現実の楽曲で出現され得る混合音のうち，ある程度はカバーできると考えられる．

4.　処理の流れ

本章では，音源同定の処理の流れについて述べる．まず，入力された音楽音響信号からスペクトログラムを求め，ピーク抽出を行う．次に，音響信号に含まれる

☆ たとえば，本稿の実験で用いる楽器音データベースには 4 楽器 1949 音が収録されているので，同時発音数が 3 の混合音は ${}_{1949}C_3$ ＝ 約 12 億通りある．これは，1 秒 1 個の速さで学習しても約 39 年かかる量である．

各単音の音高・発音時刻・音長を推定する．ただし，本稿の実験では，これらは正解を与えるものとする．その後，各単音の調波構造を抽出し，特徴抽出，次元圧縮，識別の順序で処理を進めていく．

4.1 周波数解析

入力された音楽音響信号（本稿の実験では，サンプリング周波数 44.1kHz，16 ビットリニア量子化，モノラルの音響信号を用いる）に対して，短時間フーリエ変換を用いてスペクトログラムを求める．窓関数にはハミング窓を使用し，窓幅は 8192 点，シフト長は 10ms とする．その後，フレーム毎にパワースペクトルのピークを抽出する．

4.2 単音形成

単音形成とは，スペクトルピークに含まれる調波構造を解析することにより，演奏された各単音の音高，発音時刻，音長を推定する処理をいう．ここでは，音源同定のみの性能を評価するため，各単音の音高，発音時刻，音長は正解を与えるものとする．なお，実際の応用の場面では，文献 11)，16) などの既存手法を用いることを想定している．

4.3 調波構造の抽出

前段の処理で与えられる各単音の音高，発音時刻，音長に基づいて，その単音に対応する調波構造を抽出する．音高に対応する周波数（A4 なら 440Hz，平均律で算出）の近傍 200cent 以内に存在する最もパワーの大きいピークを基本周波数成分とみなし，その周波数の整数倍のピークを 30 次倍音まで抽出する．ただし，整数倍のピークの抽出においては 5%までの誤差を許容する．その後，基本周波数の時間変化および最大パワーがそれぞれ 1 となるように周波数，パワーを正規化する．

4.4 音長の調節

音長の調節とは，特徴量テンプレート作成時に用いた単音と認識対象の単音との間で音長を合わせることにより，特徴量の音長依存性を回避することを指す．音長を合わせる最も単純な方法は，最も短い音長に合わせることである．しかし，音長を短くするとスペクトルが安定しにくくなるため，音長が長いままの状態に比べて同定が難しくなる．そこで，本稿では，いくつかの音長パターンに対して特徴量テンプレートを作成し，認識対象音の音長に応じて最も適切なものを選択する．具体的には，400ms，600ms，800ms，1000ms の 4 つの音長パターン（左からパターン I, II, III, IV と名付ける）に対して特徴量テンプレートを作成し，認識対象音より短い範囲で最長の音長パターンに合わせる．

4.5 特徴抽出

前段の処理で得られる各単音の調波構造から，楽器名の同定に効果的と考えられる特徴量を抽出する．特徴量は，我々が以前提案したもの[6]から混合音からの抽出が困難と思われるものを除いた最大 79 個（音長が 1000ms の場合．音長が 400ms，600ms，800ms のときはそれぞれ 55 個，63 個，71 個）である．以下に，使用する特徴量を列挙する．

(1) **スペクトルの時間平均に対する特徴**

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 ($i = 2, 3, \cdots, 29$)，

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越えている時間）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越えている時間）が p%である高調波成分の個数 ($p = 10, 20, \cdots, 90$)．

(2) **パワーの時間変化に対する特徴**

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 ($t = 0.15, 0.20, \cdots, 0.95$)，

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 ($t = 0.15, 0.20, \cdots, 0.95$)．

(3) **各種変調の振幅と振動数**

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．

4.6 次元圧縮

まず，主成分分析により次元を圧縮する．累積寄与率 99%で，33〜35 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．本稿では 4 種類の楽器を扱うので，特徴空間は 3 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，主成分分析のみで同次元に圧縮するのに比べて高性能になる[6]だけでなく，**2.** で述べたように，混合音から抽出した特徴ベクトルを用いることで，特徴変動にロバストな特徴量の重み

表 1 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）	音域	バリエーション	強さ	データ数*
01	ピアノ (PF)	A0–C8	1, 2, 3	強・中・弱	792
15	バイオリン (VN)	G3–E7	1, 2, 3	強・中・弱	576
31	クラリネット (CL)	D3–F6	1, 2, 3	強・中・弱	360
33	フルート (FL)	C4–C7	1, 2	強・中・弱	221

奏法は，ノーマル奏法（記号：NO）のみを使用.
* 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.

表 2 特徴量テンプレート作成に用いたデータ名と単音数

楽器	使用データ	作成された単音数*
PF	011PFNO{F,P},	[単] 487, 473, 462, 451
	01{2,3}PFNO{F,M,P}	[混 1] 41207, 23610, 21637, 15512
		[混 2] 10244, 6035, 5496, 4030
VN	151VNNO{F,P},	[単] 479, 478, 475, 474
	15{2,3}VNNO{F,M,P}	[混 1] 15880, 8863, 7998, 5627
		[混 2] 4188, 2503, 2295, 1718
CL	311CLNO{F,P},	[単] 316, 316, 315, 314
	31{2,3}CLNO{F,M,P}	[混 1] 15922, 8792, 7944, 5551
		[混 2] 4085, 2365, 2162, 1579
FL	331FLNO{F,P},	[単] 160, 160, 160, 160
	312FLNO{F,M,P}	[混 1] 10732, 5945, 5338, 3701
		[混 2] 2788, 1602, 1450, 1032

* 単音数は，左から音長パターン I〜IV のときを表す.
また，[単]，[混 1]，[混 2] は，それぞれ単一音テンプレート，混合音テンプレート 1，混合音テンプレート 2 を表す.

が大きくなるように次元を圧縮する効果がある.

4.7 識　　別

識別は，我々が以前提案した F0 依存多次元正規分布[6]を用いて行う．これは，多次元正規分布を基本周波数の関数として拡張することで，特徴量の F0 依存性を対処する手法である.

5. 評 価 実 験

提案手法の有効性を以下の評価実験で確認する.

5.1 使用データベース

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：楽器音（RWC-MDB-I-2001）[17]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.

本稿では，このデータベースからピアノ (PF)，バイオリン (VN)，クラリネット (CL)，フルート (FL) の 4 種類の楽器の音響信号を抜粋して使用する．奏法はノーマル奏法のみとする．本実験で使用するデータベースの詳細を**表 1** に記す.

5.2 評価用データ

評価用の楽曲として「蛍の光」（三重奏；約 1 分）を用いる．まず，文献 18) の楽譜を市販のシーケンサで入力し，スタンダード MIDI ファイル (SMF) として保存した．そして，SMF にしたがって RWC-MDB-I-2001 の音響信号を切り貼りするプログラムを作成し，このプログラムを用いて蛍の光の音響信号を得た．切り貼りの源となる音響信号として，011PFNOM, 151VNNOM, 311CLNOM, 331FLNOM を用いた．楽器の組み合わせに関しては，低音パートは音域の制約からピアノのみを用いたが，それ以外の制約は設けずに重複を許した．その結果，組み合わせ数は 16 通りとなった.

5.3 単一音テンプレートの作成

混合音テンプレートの作成に先立ち，単一音から特徴量テンプレートを作成する．これは，混合音テンプレートとの性能比較に用いられるだけでなく，混合音テンプレートの一部としても用いられる.

単一音テンプレートは,「蛍の光」作成には用いなかった音響信号，すなわち，バリエーション番号「1」，強度「中」（ピアノなら 011PFNOM）以外の音響信号を用いて，**4.** で述べた処理を行って作成した．この特徴量テンプレートを**単一音テンプレート**と呼ぶ．なお，特徴量テンプレート（後述の混合音テンプレートも同様）は，**4.4** で述べたように 4 つの音長パターン用のものをそれぞれ作成する.

5.4 混合音テンプレートの作成

混合音テンプレートの作成には，RWC 研究用音楽データベース：クラシック音楽（RWC-MDB-C-2001）[17]に付属する SMF を利用した．同データベースの楽曲 No. 16「クラリネット五重奏曲 イ長調 K.581 第 1 楽章」（W. A. モーツァルト；約 10 分）の SMF から第 1 バイオリン，第 2 バイオリン，ビオラの 3 パートのみを残し，評価用データの作成に用いたものと同じプログラムを利用して，三重奏の音響信号を作成した．単一音テンプレートと同様に，バリエーション番号「1」，強度「中」のもの以外の音響信号を用いた．楽器の組み合わせは,「蛍の光」と同様である.

こうして得られた三重奏の音響信号に対して，**4.** で述べた処理を行うことにより，特徴ベクトルの集合を得る．各特徴ベクトルの正解ラベル（楽器名ラベル）は，音響信号作成時に用いた SMF に基づいて付与する．そして，これらの特徴ベクトルを単一音テンプレートに追加して得られたものを混合音テンプレートとする．ここでは，**混合音テンプレート 1** と呼ぶ.

テンプレートに存在しない楽器の組み合わせが入力

表 3 実験結果

(a) 単一音テンプレート使用時

	PF	VN	CL	FL	計
PF	1891	411	21	78	2401
VN	52	376	1	39	468
CL	85	81	184	118	468
FL	120	25	24	299	468

全体 2750 / 3805 **(72.3%)**

(b) 混合音テンプレート 1 使用時

	PF	VN	CL	FL	計
PF	2071	80	71	179	2401
VN	20	319	25	104	468
CL	16	13	365	74	468
FL	101	5	43	319	468

全体 3074 / 3805 **(80.8%)**

(c) 混合音テンプレート 2 使用時

	PF	VN	CL	FL	計
PF	2053	108	60	180	2401
VN	20	310	25	113	468
CL	10	18	365	75	468
FL	96	4	46	322	468

全体 3050 / 3805 **(80.2%)**

音に出現したときに，どの程度の性能を示すかを確かめるため，楽器の組み合わせパターンを極端に減らしたものも用意した．具体的には，単一音テンプレートに，「PF–PF–PF」，「VN–VN–PF」，「CL–CL–PF」，「FL–FL–PF」の 4 種類の組み合わせのみの音響信号から得られた特徴ベクトルを追加したものを作成した．これを**混合音テンプレート 2** と呼ぶ．

表 2 に，これら 3 つの特徴量テンプレートに含まれる単音数を示す．

5.5 単一音データベースによる予備実験

多重奏の実験に先立ち，単一音に対して音源同定を行う．単一音テンプレートで学習を行い，011PFNOM, 151VNNOM, 311FLNOM, 331FLNOM に含まれる各単音について楽器名を同定する．結果は以下の通りである．ピアノ：68/68=100%，バイオリン：58/64=90.6%，クラリネット：37/40=92.5%，フルート：21/37=56.8%．フルートの認識率があまりよくなかったが，これは通常はあまり使われないような非常に高い音域において，クラリネットとして誤認識する誤りがほとんどであった．また，ピアノの分母が 66 なのは，011PFNOM に含まれる 88 音のうち，20 音は音長が 400ms 未満のために認識対象外となったからである．

5.6 三重奏の実験結果

以上の要領で作成したデータを用いて，音源同定実験を行う．単一音テンプレート，混合音テンプレート 1，混合音テンプレート 2 のそれぞれで学習し，「蛍の光」の音源同定を行って結果を比較する．音源同定の性能は，

表 4 次元圧縮後の特徴空間の各軸に対応する主な特徴量と重み値

		主な特徴量と重み値
単一音	第 1 軸	2 (-0.2485), 32 (-0.2616), 33 (-0.2422), 41 (-0.3312), 59 (0.2693), 62 (-0.2814), 63 (-0.3914), 76 (-0.2651)
	第 2 軸	3 (-0.2023), 6 (0.2191), 31 (0.6355), 40 (0.2768), 63 (0.2014), 78 (-0.2357), 79 (-0.2051)
	第 3 軸	2 (-0.2768), 3 (-0.2697), 31 (0.3163), 40 (0.3075), 76 (-0.4307), 78 (0.2947), 79 (0.3584)
混合音	第 1 軸	3 (0.3215), 31 (-0.5038), 35 (-0.2460), 41 (-0.3102), 61 (-0.2120), 62 (-0.2328)
	第 2 軸	31 (0.4961), 35 (-0.2062), 40 (0.5094)
	第 3 軸	5 (-0.4652), 12 (0.2975), 13 (-0.2657), 15 (0.2059), 31 (0.2833), 76 (-0.2345)

$$\text{認識率 } R = \frac{\text{(楽器名が正しく出力された単音数)}}{\text{(出力された全単音数)}}$$

で評価する．ただし，音長が 400ms 未満の単音は対象外とした．

実験結果を**表 3** に示す．単一音のみからなる特徴量テンプレートを用いるのに比べ，混合音から作成した特徴量テンプレートを用いることで，全体の認識率が約 8%改善された．

ここで特筆すべきは，混合音テンプレート 1 と混合音テンプレート 2 とで性能が変わらなかったことである．混合音テンプレート 1 のように，すべての楽器の組み合わせを網羅するには，対象楽器数を N として k 重奏を扱うなら，$O(N^k)$ のオーダのデータが必要である（ただし，本実験では低音パートの楽器を固定にしたので $O(N^2)$ であった）．それに対して，混合音テンプレート 2 では，$O(N)$ のオーダのデータしか使用していない．このことは，必ずしも楽器の組み合わせを網羅していなくても，混合音テンプレートが性能向上に有効であることを示している．

5.7 次元圧縮に関する考察

次元圧縮で得られた 3 次元特徴空間は，主にどのような特徴量から生成されたのかについて，各軸を生成する 1 次変換式における各特徴量の重み値（**表 4**）に基づいて議論する．まず，単一音テンプレートの場合と混合音テンプレートの場合をそれぞれ議論し，その後，両者を比較し，**2.** で述べたように，混合音テンプレートにおいて変動の大きな特徴量の重みが小さくなっているかどうかを確認する．なお，紙面の制約から，音長パターン I のものについてのみ議論する．

単一音テンプレートについて

第 1 軸は，41「パワー包絡線の近似直線の傾き」，63「最大パワーと発音開始から 0.35 秒後のときのパワーの比」の重みが 0.3 以上と大きかった他，59「最大パワーと発音時刻から 0.15 秒後のときのパワーの比」，62「最大パワーと発音時刻から 0.30 秒後のときのパワーの比」，76「振幅変調の振幅」など，重みの大きい特徴量の多くが，パワーや振幅変調に関するものであった．そのため，第 1 軸はパワーの時間変化を総合的に表す軸といえる．

第 2 軸は，31「奇数次の高調波成分と偶数次の高調波成分のパワーの比」の重みが 0.6355 と非常に大きかった．クラリネットは，奇数次の高調波成分のパワーが小さいという性質を持つため，この軸は，クラリネットとそれ以外を識別するための軸といえる．

第 3 軸は，76〜79「振幅変調・周波数変調それぞれの振幅・振動数」の重みが 0.2947〜0.4307 と大きかった．本実験で扱った 4 つの楽器のなかでは，バイオリンは，ビブラートを比較的大きめにかける楽器であり，周波数変調に特徴が現れる．また，バイオリンは，この 4 つの楽器では高調波成分がかなり高い次数まで含まれている．そのため，2「全高調波成分中の基音成分のパワーの割合」，3「全高調波成分中の 2 次までの高調波成分のパワーの割合」などのスペクトルに関する特徴（音の甲高さを表す特徴）の重みも大きくなった．

混合音テンプレートについて

第 1 軸は，31「奇数次の高調波成分と偶数次の高調波成分のパワーの比」の重みが 0.5038 と最も大きかった．この特徴量は，上で述べたようにクラリネット特有の性質を表すものである．その他，3「全高調波成分中の 2 次までの高調波成分のパワーの割合」などのスペクトルに関する特徴，41「パワーの包絡線の近似直線の傾き」などのパワーの時間変化に関する特徴についても大きな重みが得られた．

第 2 軸でも，31「奇数次の高調波成分と偶数次の高調波成分のパワーの比」の重みが 0.4961 と大きかった．これは，第 1 軸ではこれ以外にもさまざまな特徴量が反映されているために，これだけでは十分にクラリネットを識別できないからと考えられる．また，この軸では，40「音が鳴り続けている時間に対して，鳴り続けている時間が 90%以上である高調波成分の個数」の重みも大きかった．

第 3 軸では，5「全高調波成分中の 5 次までの高調波成分のパワーの割合」をはじめとして，スペクトルに関する特徴（音の甲高さを表す特徴）の重みが大きかった．

表 5 同一楽曲の異なる箇所で学習した場合の実験結果

	PF	VN	CL	FL	計
PF	2033	224	59	85	2401
VN	34	343	24	67	468
CL	15	9	376	68	468
FL	60	2	34	372	468

全体 3124 / 3805 **(82.1%)**

両者の比較

単一音テンプレートと混合音テンプレートを比較すると，単一音テンプレートでは比較的大きな重みが得られていたにもかかわらず，混合音テンプレートでは大きな重みでは現れなかった特徴量がいくつかある．これは大きく 2 つのグループに分けることができる．1 つは，パワーの時間変化に関する特徴（59，62，63など），もう 1 つは，周波数変調に関する特徴(78，79)である．これらはいずれも，周波数成分の重なりに伴う特徴変動の結果，これらの特徴量のクラス内分散・クラス間分散比（クラス間の分離度に相当）が，低下したからである．実際，パワーに関する特徴は，周波数成分が重なると壊れやすく，周波数変調は，基本周波数成分が他の音の高調波成分と重なると，正常には抽出できない．よって，**2.** で述べたように，混合音から抽出したデータを用いて線形判別分析で次元圧縮することによって，特徴変動にロバストな特徴量を重視できることを確認した．

5.8 同一楽曲内での学習・認識

混合音テンプレート作成に用いる楽曲が，認識対象の楽曲に十分に類似しているときに，どの程度の性能が得られるかを評価するため，同一楽曲内で学習・認識を行った．具体的には,「蛍の光」を前半と後半とに分け，(1) 後半で混合音テンプレートを作成して前半で認識実験，(2) 前半で混合音テンプレートを作成して後半で認識実験，を行って両者の出力結果を合わせた．**表 5** に実験結果を示す．実験の結果，全体で 82.1%の認識率が得られた．表 3(b), (c) との差は 2%未満で，表 3(b), (c) の結果は，異なる楽曲で混合音テンプレートを作成したことを考慮すると，高い性能が得られていることがわかる．

6. おわりに

本稿では，多重奏を対象とした音源同定について検討した．多重奏の音源同定における課題である，音の重なりによる特徴変動に対処するため，学習に用いるテンプレートを混合音から作成することを試みた．4 種類の楽器を対象に三重奏の認識実験を行ったところ，単一音のみから作成したテンプレートを使用したとき

の認識率 72.3%が，80.8%まで改善されることがわかった．さらに，テンプレート中の単音数を約 1/4 まで減らしても，ほぼ同等の認識率（80.2%）が得られることもわかった．

本稿の実験では，単音形成部には正解を与えたが，実際には単音形成の段階での誤りは避けることができない．そのため，後段の特徴抽出部や識別部において，単音形成に誤りが存在することを前提とした処理が求められる．今後は，こういった処理の検討を中心に研究を進めていきたい．

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（A）第 15200015 号，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),「同：クラシック音楽」(RWC-MDB-C-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），中谷 智広氏（NTT コミュニケーション科学基礎研究所），柏野 邦夫氏（同），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) Martin, K. D.: *Sound-Source Recognition: A Theory and Computational Model*, PhD Thesis, MIT (1999).
2) Eronen, A. and Klapuri, A.: Musical Instrument Recognition using Cepstral Coeddicients and Temporal Features, *Proc. ICASSP*, pp. 735–756 (2000).
3) Fraser, A. and Fujinaga, I.: Toward Real-time Recognition of Acoustic Musical Instruments, *Proc. ICMC*, pp. 175–177 (1999).
4) Fujinaga, I. and MacMillan, K.: Realtime Recognition of Orchestral Instruments, *Proc. ICMC*, pp. 141–143 (2000).
5) Agostini, G., Longari, M. and Pollastri, E.: Musical Instrument Timbres Classification with Spectral Features, *EURASIP J. Applied Signal Process.*, **2003**, 1, pp. 5–14 (2003).
6) 北原鉄朗, 後藤真孝, 奥乃博: 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, 情処学論, **44**, 10, pp. 2448–2458 (2003).
7) Marques, J. and Moreno, P. J.: A Study of Musical Instrument Classification using Gaussian Mixture Models and Support Vector Machines, *CRL Technical Report Series*, CRL/4 (1999).
8) Brown, J. C.: Computer Identification of Musical Instruments using Pattern Recognition with Cepstral Coefficients as Features, *J. Acoust. Soc. Am.*, **103**, 3, pp. 1933–1941 (1999).
9) Brown, J. C.: Feature Dependence in the Automatic Identification of Musical Woodwind Instruments, *J. Acoust. Soc. Am.*, **109**, 3, pp. 1064–1072 (2001).
10) Krishna, A. G. and Sreenivas, T. V.: Music Instrument Recognition: From Isolated Notes to Solo Phrases, *Proc. ICASSP*, **IV**, pp. 265–268 (2004).
11) 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).
12) 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).
13) 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).
14) Eggink, J. and Brown, G. J.: A Missing Feature Approach to Instrument Identification in Polyphonic Music, *Proc. ICASSP*, **V**, pp. 553–556 (2003).
15) 柏野邦夫, 村瀬洋: 単音連鎖確率ネットワークに基づく音楽演奏の音源同定, 人工知能学会誌, **13**, 6, pp. 962–970 (1998).
16) 桜庭洋平, 奥乃博: 自動採譜におけるパート形成処理のための特徴量の検討, 情処研報, 2003-MUS-51, pp. 35–42 (2003).
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：研究目的で利用可能な著作権処理済み楽曲・楽器音データベース, 情処学論, **45**, 3, pp. 728–738 (2004).
18) 柏野邦夫: 音楽音響信号を対象とする聴覚的情景分析に関する研究, 博士論文, 東京大学 (1994).