3G-4

多重奏の音源同定のための混合音からのテンプレート作成法

北原 鉄朗† 後藤 真孝‡ 駒谷 和範† 尾形 哲也† 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

1. はじめに

楽器音の音源同定は，自動採譜や音楽情報検索などにおいて重要なタスクである．しかし，これまでの音源同定研究の多くは単一音を対象としており（e.g. 1)），多重奏への取り組みが始まったのは最近である[2)~4)]．

多重奏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，これまでの研究では，波形テンプレートの適応・マッチング[2)]，特徴量の再計算[3)]，Missing Feature Theory[4)]などさまざまな対策がとられてきたが,「単一音のテンプレート（学習データ）を用いて混合音を認識する」という枠組みは共通であった．

本研究では，この問題を，混合音から作成されたテンプレートを用いて認識することで解決する．特徴変動がすでに起きているデータで学習することで，単一音のみで学習するよりロバストな同定ができると期待される．

本稿ではさらに，音楽的文脈（前後関係）に基づいて音源同定の性能を改善する方法として，2 段階による事後確率計算を検討する．各単音の事後確率を文脈を考慮せずに計算した後，前後の単音の事後確率に基づいて再計算することで，バイオリンのメロディの流れのなかで 1 音だけクラリネットが現れるといった，音楽的に不自然な誤認識を回避する．

2. 混合音からの特徴量テンプレートの作成

本稿では，周波数成分の重なりに伴う特徴変動の問題を解決するため，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．これにより次の 2 つの効果を期待できる．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．

- **学習時と認識時とで類似した特徴変動**
  特徴量テンプレートを混合音から作成することで，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，特徴変動が起きているデータに対してロバストな同定を実現できる．
- **変動の大きな特徴量への低い重みの設定**
  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とき，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り，その結果，クラス内分散・クラス間分散比が低下する．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．

しかし，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．そこで本研究では，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．

3. 音楽的文脈を考慮した事後確率計算

各単音の楽器名同定の精度向上のため，その前後単音の情報（音楽的文脈）を利用する．文脈利用の基本的アイディアは，単音 n_kに対する事後確率 $p(\omega_i|\boldsymbol{x}_k) = p(\boldsymbol{x}_k|\omega_i)p(\omega_i)/p(\boldsymbol{x}_k)$ の計算において，事前確率 $p(\omega_i)$ に前後の単音に対する事後確率を利用することである．ここで$\boldsymbol{x}_k$は単音 n_kから観測された特徴ベクトル，ω_iは楽器番号である．これを以下の 2 段階処理で実現する．

[第 1 パス] 文脈を考慮しない事後確率の仮計算

各単音に対して事前確率を定数として事後確率を計算する．事後確率計算までの処理の流れは後述する．

[第 2 パス] 文脈を考慮した事後確率の再計算

各単音 n_kに対して，以下の処理を行う．

(1) 文脈的に単音 n_kと同じ楽器で演奏されたと判断できる単音を発音時刻が n_kに近いものから前後各 c 個抽出する．本稿では，2 単音 n_kと n_jが同じ楽器によるものかを，各パートの音高が交差することは少ない[5)]ことに着目し，高い方から（低い方から）何番めの音かに基づいて判定する．単音 n_kの発音中に n_kよりも高い音域で発音する単音の最大同時発音数を $s_{\mathrm{h}}(n_k)$，低い音域で発音する単音の最大同時発音数を $s_{\mathrm{l}}(n_k)$ とすると，$s_{\mathrm{h}}(n_k) = s_{\mathrm{h}}(n_j)$ かつ $s_{\mathrm{l}}(n_k) = s_{\mathrm{l}}(n_j)$ のとき，n_kと n_jは同一パート（同一楽器による演奏）とみなす．以下，抽出された単音の集合を $\mathcal{N}$ で表す．

(2) 前後関係から単音 n_kが楽器ω_iと判断できる確率 $p(Z_{n_k} = \omega_i)$ を求める．ここで，Z_{n_k}は単音 n_kの楽器名を表す確率変数である．これは，

$$p(Z_{n_k} = \omega_i) = p(Z_{n_k} = \omega_i \mid \forall n_j \in \mathcal{N} : Z_{n_j} = \omega_i) \times \prod_{n_j \in \mathcal{N}} p(Z_{n_j} = \omega_i)$$

と変形できる．この右辺の第一因子は統計的分析によって得ることもできるが，ここでは簡単のため $1-(1/2)^{2c}$ を用いた．この式は，考慮する前後の単音数が多いほど，そこから得られる情報の信頼性が高いことを表現したものである．また，$p(Z_{n_j} = \omega_i)$ は第 1 パスで計算した事後確率を用いる．

(3) 上の方法で求めた $p(Z_{n_k} = \omega_i)$ を事前確率として，単音 n_kの事後確率を再計算する．再計算後，事後確率が最大となる楽器名を同定結果と決定する．

Feature Template Construction from Sound Mixtures for Instrument Identification in Polyphonic Music
Tetsuro Kitahara†, Masataka Goto‡, Kazunori Komatani†, Tetsuya Ogata† and Hiroshi G. Okuno†
(†Kyoto Univ., ‡Nat'l Inst. of Adv. Ind. Sci. and Tech.)

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.
バリエーション「1」，強さ「中」のデータを評価用に，その他をテンプレート作成用に割り当てる.
* 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.

4. 事後確率計算の処理の流れ

事後確率計算までの処理の流れは以下の通りである.
(1) 入力された音楽音響信号に対して，短時間フーリエ変換を用いてスペクトログラムを求め，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.
(2) 各単音の音高・発音時刻・音長を推定する．ただし，本稿では音源同定のみの性能を評価するため，正解を与える.
(3) 推定された音高に基づいて各単音の基本周波数成分と高調波成分（10 次まで）のピークを抽出する．その後，単音毎に，基本周波数の時間平均，最大パワーがそれぞれ 1 になるように正規化する.
(4) 特徴量の音長依存性を回避するため，認識対象音の音長をテンプレート作成に用いた音長に合わせて短くする．テンプレートは 300ms，450ms，600ms の 3 パターンで作成し，単音毎に当該音より短い範囲で最長の音長パターンが選ばれる．なお，300ms 未満の単音は同定の対象外とする.
(5) 各単音の調波構造から,「周波数重心」「パワー包絡線の近似直線の傾き」など，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なものを除いた最大 43 個（音長パターンに依存）の特徴量を抽出する.
(6) 主成分分析で 21 次元（累積寄与率 99%）に圧縮したのち，線形判別分析でさらに次元を圧縮する．ここでは 4 楽器を扱うので 3 次元となる．これにより，特徴変動が大きな特徴量の重みが小さくなり，変動にロバストな特徴空間が構成される.
(7) 上により得られた 3 次元特徴空間上で特徴ベクトルが F0 依存多次元正規分布[6]に従うと仮定し，ベイズ決定規則により事後確率を計算する.

5. 評 価 実 験

RWC 研究用音楽データベース（楽器音）[7]の音響信号（**表 1**）をスタンダード MIDI ファイル（SMF）に従って切り貼りして作成した三重奏および二重奏の音響信号に対して同定実験を行った．SMF には RWC 研究用音楽データベース（クラシック）[7]の No. 13, 16, 17 から 3 あるいは 2 パートを抜粋して使用した．混合音テンプレートは，認識対象曲以外の 2 曲を用いて作成した．実験結果を**表 2** に示す．混合音からの特徴量テンプレート作成および音楽的文脈の利用により，平均の認識率が，三重奏で 58.2%から 83.6%まで，二重奏で 62.2%から 88.9%まで改善された．特に CL，FL において，34〜58%から 71〜95%へと認識率が大幅に改善された．また，次元圧縮においては，パワーの時間変化や振幅変調など，音の混合で変動しやすい因子の負荷量が低くなることが確認された.

表 2 実験結果

	テンプレート 音楽的文脈		単一音 なし	単一音 あり	混合音 なし	混合音 あり
三重奏	No. 13	PF	82.0%	85.2%	88.6%	94.2%
		VN	62.4%	79.6%	69.4%	84.9%
		CL	42.9%	36.9%	70.6%	81.6%
		FL	46.2%	63.0%	73.3%	78.9%
	No. 16	PF	89.8%	94.7%	91.9%	97.9%
		VN	55.6%	71.8%	55.1%	79.5%
		CL	47.7%	42.3%	80.2%	90.8%
		FL	57.4%	70.9%	66.4%	80.4%
	No. 17	PF	81.2%	85.4%	84.4%	89.4%
		VN	51.8%	72.6%	60.1%	77.8%
		CL	34.2%	26.8%	62.7%	76.9%
		FL	46.9%	53.3%	69.1%	71.5%
	平均		58.2%	65.2%	72.6%	83.6%
二重奏	No. 13	PF	92.4%	94.3%	94.8%	98.3%
		VN	61.3%	79.4%	66.5%	85.6%
		CL	57.4%	61.7%	83.0%	92.6%
		FL	39.6%	53.5%	72.3%	89.1%
	No. 16	PF	95.8%	98.2%	96.5%	99.0%
		VN	58.2%	76.3%	54.8%	75.3%
		CL	58.2%	66.0%	83.7%	94.8%
		FL	53.9%	71.6%	57.4%	80.9%
	No. 17	PF	91.5%	94.2%	92.6%	96.6%
		VN	58.8%	85.3%	60.9%	85.6%
		CL	42.5%	45.1%	73.2%	92.8%
		FL	36.5%	52.1%	66.5%	76.0%
	平均		62.2%	73.1%	75.2%	88.9%

6. お わ り に

本稿では，高精度な多重奏の音源同定を実現するため，混合音からの特徴量テンプレート作成および音楽的文脈の利用について検討し，実験により認識率の改善を確認した.

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)，特定領域「情報学」）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」の支援を受けた.

参 考 文 献

1) K. D. Martin: *Sound-Source Recognition: A Theory and Computional Model*, PhD Thesis, MIT, 1999.
2) 柏野 他：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.
3) 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.
4) J. Eggink *et al.*: A Missing Feature Approach to Instrument Identification in Polyphonic Music, *Proc. ICASSP*, **V**, pp.553–556, 2003.
5) Y. Sakuraba *et al.*: Comparing Features for Forming Music Streams in Automatic Music Transcription, *Proc. ICASSP*, **IV**, pp.273–276, 2004.
6) 北原 他：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) 後藤 他：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.