

多重奏中特定パートの自動採譜における 複数特徴量の自動重み付け

糸山 克寿[†] 北原 鉄朗[‡] 駒谷 和範[‡] 尾形 哲也[‡] 奥乃 博[‡]

[†] 京都大学 工学部情報学科

[‡] 京都大学大学院 情報学研究科 知能情報学専攻

1. はじめに

ピアノなどの単一楽器による演奏の自動採譜は実用の域に入りつつあるが、複数楽器による演奏（多重奏）の自動採譜は、いまだ困難な課題である。多重奏の自動採譜では、各単音の音高や発音時刻を推定する処理（単音形成）の後に、形成された単音をパート（楽器）ごとに分類する処理（パート形成）が必要となる。従来研究において後者を扱ったものは前者に比べて少なく、多重奏の自動採譜が難しい一つの要因となっている。

これまでのパート形成研究の問題設定は、2つに大別できる。1つは、対象楽曲に含まれる全楽器を既知とした上で、形成されたすべてのパートの楽器を同定して、楽器ラベルを付与するものである [1, 2, 3]。この問題設定では、対象楽曲に含まれ得るすべての楽器に関して学習データを用意する必要がある。

もう1つは、対象楽曲に含まれる楽器情報は与えずにパート形成を行う。この問題設定では、形成されたパートに楽器ラベルが一切付与されないため、楽器同定のための学習データが不要である [4, 5]。しかし、採譜結果中の特定の楽器パートに着目することはできない。

本研究の問題設定は、指定された1つの楽器に着目してその楽器パートのみを形成するというもので、両者の中間に位置する。本問題設定では、基本的に指定楽器の学習データさえ用意すれば、特定の楽器パートの自動採譜が可能になると期待される。本問題設定に基づくパート形成を「特定パート追跡」、自動採譜処理全体を「特定パートの自動採譜」と呼ぶ。

本稿ではさらに、特定パート追跡の性能向上のために特徴量の自動重み付けについても検討する。特定パート追跡では、音色の類似性や定位の近接性、音高遷移など、複数の手がかりを統合して処理を進める。このような手がかりの重要度は採譜すべき音響信号の録音条件などによって異なる。そのため、音響信号ごとに準最適な重みを自動的に設定する方法についても述べる。

2. 特定パート自動採譜におけるパート追跡

本章では、特定パートの自動採譜のための、特定パート追跡処理について述べる。

2.1 パート追跡処理の概要

特定パート追跡処理は、単音形成処理で得られた単音列 $n_1 n_2 \dots$ を、指定された楽器（対象楽器）に由来するもの N とそうでないもの \bar{N} に分類する処理である。 N と \bar{N} の組 $\mathcal{H} = (N, \bar{N})$ をパート追跡の仮説とすると、次のような仮説探索で解くことができる。

1. 最初の単音 n_1 に対して、2つの初期仮説 $(\{n_1\}, \phi)$, $(\phi, \{n_1\})$ を生成する。
2. 単音列 $n_1 \dots n_k$ について得られたパート追跡の各仮説について、仮説の展開を行う。展開すべき仮説

を $\mathcal{H} = (N, \bar{N})$ とすると、新たな2つの仮説 $\mathcal{H}_0 = (N \cup \{n_{k+1}\}, \bar{N})$ と $\mathcal{H}_1 = (N, \bar{N} \cup \{n_{k+1}\})$ を生成し、それぞれの評価値 $f(\mathcal{H}_0), f(\mathcal{H}_1)$ （仮説の尤度に相当する）を計算する。なお、仮説数が一定数 K を超えた場合は、評価値が上位 K 個の仮説のみを残し、それ以外の仮説は消去する。

3. すべての単音に対して、2を繰り返す。単音列の最後まで仮説の展開と評価を行い、評価値が最大の仮説をパート追跡結果とする。

仮説の評価値 $f(\mathcal{H})$ は、複数の手がかり（特徴量）に基づいて以下のように評価する。 $f_i(\mathcal{H})$ は各特徴量の評価値、 w_i は特徴量への重みである。

$$f(\mathcal{H}) = \sum_i w_i f_i(\mathcal{H}) \quad (1)$$

2.2 評価値計算法

評価値の計算は、既存の仮説に単音が追加されたときに行われる。仮説の評価値は各単音に対する評価値の平均で表す。追加された単音を n 、既存の仮説のうち対象楽器側の単音集合を N とすると、評価値を計算する特徴量は、 n の対象楽器らしさを表すものと、 n と N を形成する単音との同一パートらしさを表すものに分けられる。本研究では、前者には「楽器単音リファレンスモデルとの音色類似性」を用い、後者には「音色類似性」「方向近接性」「音高遷移頻度」を用いる。以下で各特徴量について述べる（紙面の制約上、詳細は省略する）。なお、後者の3つは桜庭らの手法 [5] を参考に設計した。

楽器単音リファレンスモデルとの音色類似性 各単音の音色を、周波数重心や倍音成分のパワー比などの、北原らが提案した特徴ベクトルで表現する [6]。単音 n の音色特徴ベクトルと特定楽器の単音に関するリファレンスモデルとのマハラノビス汎距離を、楽器音一般を表現した「グローバルな」単音モデルとの距離で正規化し、F分布の片側検定によって0から1で表される評価値を算出する。

音色類似性 n と N を形成する単音の音色の類似性を0から1で表現したものである。音色は、上と同様に北原らの特徴ベクトル [6] を用いて表現し、 χ^2 分布の片側検定によって評価値を求める。

方向近接性 n と N を形成する単音の方向の近接性を0から1で表現したものである。方向は、フレーム毎に求めたIPD列を1度間隔でヒストグラムをつくったときの最頻値とする。

音高遷移頻度 N の後に n が続く確率を trigram 確率で評価する。

これらのうち、「音色類似性」と「方向近接性」は、 N 内の全単音の音色ベクトルや方向の平均や分散を用いるため、 N の要素数が少ないときには信頼性が低くなる。従って、次式により信頼度を考慮した評価値計算を行う。

$$f'_i(n) = f_i(n) - 1.96 \sqrt{f_i(n)(1 - f_i(n)) / |N|} \quad (2)$$

Automatic feature weighting in automatic transcription of specified part in polyphonic music: Katsutoshi Itoyama, Tetsuro Kitahara, Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno (Kyoto Univ.)

3. 複数特徴量の自動重み付け

上記の評価値を求める上での問題は、パート追跡に最適な重みは認識対象の音響信号の録音条件などによって異なるので、音響信号から自動的に求める必要があることである。各音響信号に最適な重みを自動で決定するには、パート追跡の「よさ」を表す評価基準を設け、これを最大化する重みを探索する。 $\mathcal{H} = (N, \bar{N})$ がよいパート追跡結果である条件は、次の2条件で規定できる。

1. N において、対象楽器に由来する単音が \bar{N} よりも十分に優勢である。
2. N に含まれる単音の多くが同一の音源に由来する。各々に対応する定量的な評価法を次のように設計した。

1. 単音リファレンスモデルに関する特徴量の評価値 $f_I(\mathcal{H})$ の N と \bar{N} の間の平均の差

$$E[f_I(N)] - E[f_I(\bar{N})] \quad (3)$$

2. 各特徴量に関する評価値のクラス内分散・クラス間分散比の和

$$\sum_i \frac{(E[f_i(N)] - E[f_i(\bar{N})])^2}{\text{Var}[f_i(N)] + \text{Var}[f_i(\bar{N})]} \quad (4)$$

上記の2つの値の積をパート追跡の「よさ」を表す評価基準とする。これを最大化する重みは理論的には求まらないので、遺伝的アルゴリズム(GA)を用いて探索する。

4. 実験

パート形成処理を2つの実験で評価する。

1. 本稿で提案する重み付け手法の総合的な性能の評価で、Vn, Fl, Pfの3つの楽器で演奏される音響信号に対して各々の楽器のパート追跡を行う。
2. 未知楽器が存在する場合にも特定パートの追跡が可能かどうかを評価するため、Vn, Fl, Pfの3つの楽器で演奏される音響信号のPfパートの追跡結果と、VnとFlをTrやClに置き換えた音響信号のPfパートの追跡結果を比較する。

実験に用いた楽曲は「蛍の光」であり、音響信号はSMFファイルをもとに、楽器単音を貼り付けて作成した。この楽曲は3つのパートから構成されており、各パートに割り当てる楽器を変えることで、複数のパターンを作成した。楽器単音にはRWC-MDB-I-2001[7]を用いた。いずれの実験でも、個体番号は1のものを用いた。

楽器単音リファレンスモデルは、RWC-MDB-C-2001に含まれる楽曲に付属する楽譜に基づいた混合音テンプレート[6]を用いて学習を行った。評価楽曲と同様に、音源はRWC-MDB-I-2001を用いている。実験1ではVn, Fl, Pfに対して、すべての楽器個体を用いて作成し、実験2ではPfに対して、楽曲作成に用いなかった個体のみを用いて作成した。また、グローバルな単音モデルは、いずれの実験でもVn, Fl, Pfに対して作成した楽器単音リファレンスモデルを1つのモデルにまとめたものを用いている。すなわち、TrとClはグローバルな単音モデルにも含まれていない未知楽器であり、これらが含まれる音響信号に特定パート追跡を行うことで、未知楽器を含む多重奏中の特定パート追跡への評価を行える。

比較した対象は、自動重み付けを行わない場合の結果と、パート追跡正解データと比較した場合のF値をGAの適応度として用いた場合のA Prioriな結果である。

表1: 実験1の結果

楽器構成	追跡	重みなし	本手法	A Priori
Vn-Fl-Pf	Vn	93.8%	80.0%	94.7%
	Fl	11.4%	52.3%	63.7%
	Pf	80.4%	87.7%	98.4%
total		64.5%	76.7%	89.2%
Fl-Vn-Pf	Fl	97.4%	92.9%	98.3%
	Vn	1.7%	74.0%	76.0%
	Pf	90.2%	81.9%	98.4%
total		69.1%	82.9%	93.5%

表2: 実験1での重みパラメータの比較

楽器構成	追跡	適応度	w_I	w_S	w_L	w_T
Vn-Fl-Pf	Vn	本手法	0.69	0.02	0.07	0.19
Vn-Fl-Pf	Vn	A Priori	0.29	0.22	0.31	0.16
Vn-Fl-Pf	Fl	本手法	0.37	0.13	0.05	0.43
Vn-Fl-Pf	Fl	A Priori	0.32	0.16	0.03	0.43

表3: 実験2の結果

楽器構成	追跡	重みなし	本手法	A Priori
Vn-Fl-Pf	Pf	90.7%	82.8%	94.0%
Vn-Cl-Pf	Pf	92.1%	99.5%	99.5%
Tr-Fl-Pf	Pf	70.2%	85.0%	85.0%
Tr-Cl-Pf	Pf	71.5%	99.5%	99.5%

実験結果を表1と表3に示す。数値は正解データと比較したときのF値(再現率と適合率の調和平均)である。

実験1では、全楽器パートの結果を結合すると、F値が13%程度上昇した。本手法での重みとA Prioriな場合の重みを比較する(表2)と、F値がA Prioriな場合に近いパート追跡結果では、重みもA Prioriな場合に近くなることも確認できた。しかし、F値が低下しているパートもあり、パート追跡の「よさ」における評価基準を改善する必要がある。

実験2では、自動重み付けを行わない場合は未知楽器数の増加に伴ってF値は減少する。一方、本手法で重み付けを行うと未知楽器数が増加してもF値を維持できおり、「音響信号に含まれるすべての楽器が分からなくとも、特定楽器パートを追跡できる」ことを示している。

5. おわりに

本稿では、多重奏中の特定パート追跡という問題設定において、複数特徴量を統合するために自動的に重み付けを行う手法について述べ、未知楽器を含む音響信号に対しても、特定楽器パートの追跡が可能であることを示した。今後はより複雑な楽曲に対しても本手法を適用し、有効性を確かめていきたい。

謝辞 本研究の一部は、日本学術振興会科学研究費補助金、21世紀COEプログラムの支援をうけた。

参考文献

- [1] 柏野他: “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-D-II, No.11, pp.1751-1761, 1996.
- [2] 柏野他: “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-D-II, No.11, pp.1762-1770, 1996.
- [3] 柏野他: “単音連繫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 Vol.13, No.6, pp.962-970.
- [4] 木下他: “自動採譜処理における知覚的階層に着目したパート分離処理”, 信学論 D-II Vol. J85-D-II, No.3, pp.382-389, 2002.
- [5] Y. Sakuraba, et al.: “Comparing Features for Forming Music Streams in Automatic Music Transcription”, ICASSP 2004, Vol. IV, 273-276.
- [6] T. Kitahara, et al.: “Instrument Identification in Polyphonic Music: Feature Weighting with Mixed Sounds, Pitch-dependent Timbre Modeling, and Use of Musical Context”, ISMIR 2005, pp.558-563.
- [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, Vol.45, No.3, pp.728-738, 2004.