

## 自動採譜における音色と定位と音楽知識を統合したパート形成

櫻庭 洋平 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

## 1. はじめに

自動採譜は、作編曲者の補助となるだけでなく、音楽検索などに応用できる重要な技術である。複数楽器による多重奏の自動採譜には、入力した音楽音響信号から得られた単音列をパート（同一楽器に演奏された単音の集合）ごとに分類するパート形成処理が必要である。本稿では、自動採譜システムのパート形成処理に焦点を当て、この処理に有効な特徴量について議論する。

一般に、パート形成処理の重要な特徴量は楽器の音色である<sup>1)2)</sup>。しかし、複数音が同時に発音すると、周波数成分の干渉により、周波数成分のパワー変化などの音色特徴は正確に抽出できない。この問題に対処するために、我々は音色類似度と定位近接度に基づいてパートを形成したが、より多くの特徴量を統合するという課題が残されていた<sup>3)</sup>。

本稿では、パート形成性能を改善するために、音色類似度、定位近接度、音高遷移度、音高高低維持度の4つの手がかりを用いてパートを形成することを提案し、その有効性および、特徴量間の優位性について検証する。

## 2. パート形成処理

パート形成処理とは、音楽音響信号から得た単音列（音高、オンセット、オフセットの組とそれに対応するスペクトログラム）をパートごとに分類する処理である。中心となる処理は、パート  $s$  と単音  $n$  が同一楽器で演奏されたかどうかの判定である。このために、我々は音色類似度  $TS(s, n)$ 、定位近接度  $DP(s, n)$ 、音高遷移度  $PT(s, n)$ 、音高高低維持度  $PRC(s, n)$  の4つの特徴量を用いる。

音色類似度は、一般に用いられている特徴量である。音色は、北原らによって提案された音色特徴ベクトルで表現する<sup>4)</sup>。定位近接度は、音色類似度の曖昧性解消に有効な特徴量である<sup>3)</sup>。音高遷移度と音高高低維持度は、今回新たに導入した特徴量である。

音高遷移度により、調性音楽における音高の遷移の傾向を統合する。調性音楽では、音高の遷移はランダムに現れるわけではない。パート  $s$  が単音  $n$  とパートを形成した場合の遷移が調性音楽でよく見られるのであれば、それらは同一楽器に演奏されたと判断できる。

音高高低維持度により、パート間の音高の高低（どちらのパートがより高音を演奏するか）の傾向を統合する。一般に、パート間の音高の高低は維持される傾向にある。この傾向は、弦楽四重奏（第一バイオリン、第二バイオリン、ピオラ、チェロ）において、最高音の大部分は第一バイオリンが

演奏し、最低音の大部分はチェロが演奏することからもわかる。もし、 $s$  が  $n$  とパートを形成したとき、音高高低が維持されれば、それらが同一楽器に演奏されたと判断できる。

## 2.1 パート形成のための特徴量の設計

## 2.1.1 音色類似度

各単音の音色を、周波数重心や各倍音のパワー比などの198次元の音色特徴ベクトルで表現する<sup>4)</sup>。パートの音色特徴ベクトルは、それに属する全単音の音色特徴ベクトルの平均とする。パート  $s$  と単音  $n$  の音色間の距離は、 $s$  と  $n$  の音色特徴ベクトルのユークリッド距離と定義する。距離から確率への変換においては、パート  $s$  に属する全単音の音色の分布に関して、正規分布を仮定した。ただし、各音色特徴は分散が1となるように正規化してある。

## 2.1.2 定位近接度

単音  $n$  の定位  $D(n)$  は、2本のマイク間の位相差から求める<sup>5)</sup>。パート  $s$  の定位  $D(s)$  は、それに属する全単音の定位の平均とする。このとき、定位近接度を

$$DP(s, n) = 1 - \frac{|D(s) - D(n)|}{2 T_d(D(s))},$$

と定義する。定位の閾値  $T_d(x)$  を人間の聴覚のように、正面で分解能が高く、外側で低くなるように設計した。

$$T_d(x) = T_c + (T_o - T_c) \cdot \frac{|x|}{90}.$$

$T_c = 10$ ,  $T_o = 20$  はそれぞれ正面 (0 deg), 外側 ( $\pm 90$  deg) における方向の近接度を決定する閾値である。

## 2.1.3 音高遷移度

音高遷移度を得るために、RWC 研究用音楽データベース：クラシック音楽 (RWC-MDB-C-2001<sup>6)</sup>) の全曲 (50曲, 総音符数 167179音) を解析し、音高のトライグラムモデルを得た。このとき、音高遷移度を

$$PT(s, n) = p(n|n_{c-1}, n_c),$$

と定義する。 $n_i (1 \leq i \leq c)$  は  $s$  に属する  $i$  番目の単音、 $c$  をそれらの総数である。モデルは、長調と短調それぞれ作成した。また、音高の遷移の頻度は、曲の調性に依存するため、音高を調性により正規化してある。

## 2.1.4 音高高低維持度

音高高低維持度を得るために、同データベース<sup>6)</sup> を次のように解析した。

- (1) 各曲から2パートの組み合わせをすべて抽出する。
- (2) 各組み合わせに対し、音高の高低  $PR$  を調べる。2パートのうち、より多く高音を発音しているパートをパート  $A$ 、他方をパート  $B$  とすると、 $PR$  は次の式で求められる。

$$PR = \frac{\text{パート } A \text{ がより高音を発音している時間}}{2 \text{ パートが同時に発音している時間}}$$

- (3) 全組み合わせの平均をその曲の音高高低維持度とする。50曲の音高高低維持度の平均は0.88であった。

Forming music stream in automatic music transcription by integrating timbre, direction and musical knowledge by Yohei Sakuraba, Tetsuya Ogata and Hiroshi G. Okuno (Kyoto Univ.)

表 1 実験データ

曲名	パッヘルベルのカノン	蛍の光
演奏形態	四重奏	三重奏
演奏時間	6分30秒	1分
単音数	5868	726
楽器配置 (左から)	violin, flute, trumpet, piano	flute, violin, piano
楽器間隔	20°, 40°, 60°間隔	30°, 60°, 90°間隔

パート  $s$  が単音  $n$  とパート形成をすることで、他のパート  $s_i$  ( $1 \leq i \leq N$ ) と音高高低を維持する(維持しない)とき、音高高低維持度を次のように定義する。

$$PRC(s, n) = \frac{1}{N} \sum_i x_i$$

$$x_i = \begin{cases} 0.88 & \text{パート } s \text{ とパート } s_i \text{ が音高高低を維持する} \\ 0.12 & \text{パート } s \text{ とパート } s_i \text{ が音高高低を維持しない} \end{cases}$$

### 2.2 ビームサーチによるパート形成仮説の探索

複数のパート形成仮説(どのパートとどの単音がパートを形成するかを表す仮説)を生成し、ビームサーチにより仮説を探索する。各仮説は、処理単位(楽譜における32分音符に相当する時間)ごとにパートと単音とのパート形成の全組み合わせを生成し、以下で説明する仮説の信頼度の上位5個のみを保存する。このとき、仮説の生成において、一定時間どの単音ともパートを形成しないパートは、発音を終了したとみなし、パート形成の候補から取り除く。

現在の処理単位における、各仮説の信頼度は、それまでの処理単位のパート形成信頼度の平均とする。各処理単位におけるパート形成信頼度は、各パートのパート形成信頼度の平均とする。各パートのパート形成信頼度  $L$  は、4つの特徴量の積と定義する。

$L(s, n) = TS(s, n) \times DP(s, n) \times PT(s, n) \times PRC(s, n)$ .  
曲が終了したときに、最も信頼度の高い仮説をパート形成結果として出力する。

### 3. 評価実験

本章では、4つの特徴量を用いたパート形成実験を行う。ベンチマークデータには「パッヘルベルのカノン」と「蛍の光」(表1)を用いた。AKAIのサンプラーS6000(楽器音データベースNTTMSA-P1を保存したもの)でパートごとに異なるスピーカーを用いて同時に再生し、2本のマイク(マイク間の距離20cm)を用いて無響室で録音した。

出力パートに含まれる単音のうち、楽譜で同一音源に由来するものを正解とした。パート形成精度  $R$  は

$$R = \frac{\text{正しくパート形成された単音の個数}}{\text{楽譜に含まれる単音の個数}}$$

とした。4つの特徴量のうち、1つだけを用いた場合から4つ全てを用いた場合の合計15通りでパート形成精度を比較した。実験の結果を表2に示す。

実験の結果、4つの特徴量の中では定位近接度が最も有効であった。定位は音色と比べ、混合音における干渉の影響を

表 2 パート形成結果

音色類似度	定位近接度	音高遷移度	音高高低維持度	カノン	蛍の光
○	-	-	-	52.97%	69.83%
-	○	-	-	76.72%	78.93%
-	-	○	-	63.19%	69.56%
-	-	-	○	56.88%	75.07%
○	○	-	-	76.38%	78.79%
○	-	○	-	61.45%	70.25%
○	-	-	○	56.13%	73.00%
-	○	○	-	79.19%	80.99%
-	○	-	○	77.79%	84.71%
-	-	○	○	60.43%	79.06%
○	○	○	-	77.88%	80.58%
○	○	-	○	77.42%	86.50%
○	-	○	○	59.63%	74.52%
-	○	○	○	80.01%	84.71%
○	○	○	○	78.61%	87.05%

受けにくいと考えられる。音色類似度のみと4つの特徴量を統合した場合を比較すると、カノンでは52.97%から78.61%へ、蛍の光では69.83%から87.05%へ性能が向上した。カノンでは、音色類似度を統合することにより、性能が低下した。パート形成精度の向上のためには、混合音にも有効な音色特徴について検討する必要がある。また、曲により各手がかりの有効性が異なることがわかる。今後、各手がかりの統合の重みについてさらに検討する必要がある。

### 4. おわりに

本研究では、自動採譜におけるパート形成の特徴量として、音色類似度、定位近接度、音高遷移度、音高高低維持度の4つを提案した。4つの特徴量を統合することで、音色類似度のみと比べ約25%性能が向上した。今後は、各特徴量の重みを検討や、コード進行や類似パターンの繰返しなどの曲の構成に関する情報との統合について研究を進める予定である。

謝辞 本研究の一部は、科研費基盤研究(A)第15200015号、サウンド技術振興財団研究助成、および21世紀COEプログラムの支援を受けた。また、音響信号データNTTMSA-P1の使用許可を下されたNTTコミュニケーション科学基礎研究所、無響室を貸して下さった国際電気通信基礎技術研究所に感謝する。

### 参考文献

- 1) 柏野 邦夫, 中臺 一博, 木下 智義, 田中 英彦: 音源情景分析の処理モデル OPTIMA における単音の認識, 信学論, J79-DII, 11, pp.1751-1761, 1996.
- 2) K. Kashino and H. Murase: 単 A Sound Source Identification System for Ensemble Music Based on Template Adaptation and Music Stream Extraction, *Speech Communication*, 27, pp.337-349, 1999.
- 3) Y. Sakuraba and H. G. Okuno: Note Recognition of Polyphonic Music by Using Timbre Similarity and Direction Proximity, *Proc. ICMC*, pp.167-170, 2003
- 4) 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.
- 5) Y. Sakuraba, T. Kitahara and H. G. Okuno: Comparing Features for Forming Music Streams in Automatic Music Transcription, *Proc. ICASSP*, 2004.
- 6) 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: クラシック音楽データベースとジャズ音楽データベース", 情処研報, 音楽情報科学, 2002-MUS-44, pp.25-32, 2002