

Factorial HSMMに基づく 音楽音響信号に対するリード・リズムギター譜推定

柴田 健太郎[†] 坂東 宜昭[‡] 尾島 優太[‡] 錦見 亮[‡] 糸山 克寿[‡] 吉井 和佳[‡]

[†] 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

音楽の楽しみ方の主要なもののひとつに、ポピュラー音楽のカバー演奏があるが、バンド譜を入手することは必ずしも容易ではない。バンドでのカバー演奏は、録音された市販楽曲の楽器構成とは異なり少人数の限られたパートで演奏される。よってバンド譜は、元楽曲の主要でないパートを削減、または統合して少人数編成用に構成されている必要がある。本研究では、ポピュラー音楽をカバー演奏用バンド譜における、ギターパートを推定する問題を取り扱う。ポピュラー音楽では伴奏パートは大きく二分され、片方がメロディ要素を担当するリードパート、もう一方は伴奏要素を担当するリズムパートである。そこで、本稿では元楽曲がバンド編成でない場合であっても、バンド編成のギター2本でリード・リズムパートを再現する楽譜を推定する手法を提案する。

カバー演奏用の採譜について、Arigaら [1] はポピュラー音楽をソロギターで演奏可能な楽譜を生成する手法を提案している。この手法ではルールベースに加えて、音響信号を考慮することで原曲の聴感に近い楽譜の生成を可能にしている。本研究ではリード・リズムパートを分けて推定し、バンド演奏可能な採譜を試みる。

本研究ではポピュラー音楽のリード・リズムパートの奏法の違いに着目し、これらの統計的特徴に基づいてリード・リズムパートの採譜を行う。具体的には、リードパートは主に単音でメロディを演奏し、リズムパートはコードをシンプルなりズムの繰り返しで表現するという仮定を置く。リード・リズムギターそれぞれを隠れセミマルコフモデル (HSMM) で定式化し、それに従いリードギターとリズムギターの音響信号が出力され、それらの出力の和で推定対象楽曲の音響信号を近似する。

2. 提案手法

本章では、リズム・リードギター譜の推定手法について述べる。モデル全体は、HSMM からなる音高の遷移を表現するリードギターモデル及びコードの遷移を表現するリズムギターモデルで構成される (図1)。リードギターは音符列を生成し、音符からスペクトルを生成し、リズムギターはコード列を生成しコードからスペクトルを生成する。これらが重畳して混合スペクトログラムを生成する階乗隠れセミマルコフモデル (FHSMM) を考える。与えられた観測データに対して生成モデルの逆問題を解くことで、潜在変数であるコード列と音符列を最尤推定する。

本稿で扱うギターパート譜推定の問題設定を示す。入力は歌声分離・打楽器音分離ビート時刻解析済みの音楽音響信号の振幅スペクトログラム $\mathbf{X} \in \mathbb{R}_+^{F \times T}$ であり、出力はリズム・リードギターのパート譜である。ここで F

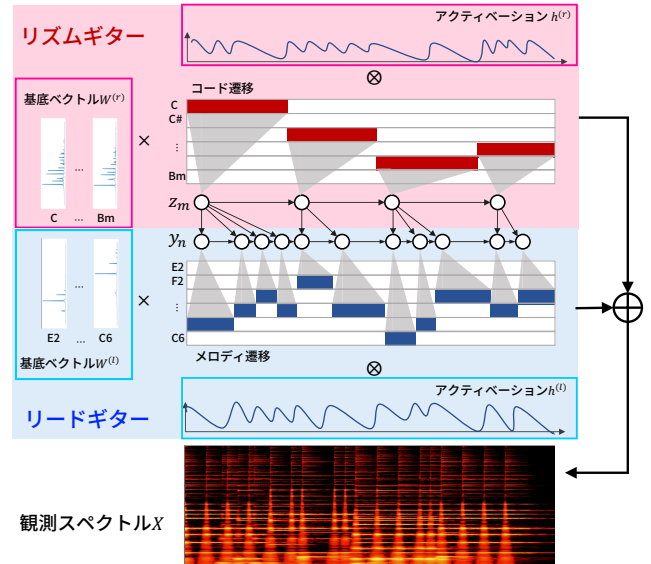


図1: モデル全体像

は周波数ビン数、 T は時間フレーム数を表す。歌声分離については Jansson ら [2] による畳み込みニューラルネットワーク (CNN) を用いた歌声分離手法を用いた。打楽器音分離として Ono ら [3] による調波・非調波分離手法 (HPSS) を用いた。ビート時刻は 16 音符単位とする。

2.1 階乗隠れセミマルコフモデル

リズムギターモデル リズムギターのコード系列の生成過程を定式化する。リズムギターの HSMM はコード遷移を潜在変数系列 $\mathbf{Z} = \{z_1, \dots, z_M\}$ ($z_m \in \{1, \dots, I\}$) に持ち振幅スペクトルの基底の系列 $\hat{\mathbf{W}}^{(r)} = \{\hat{\mathbf{w}}_1^{(r)}, \dots, \hat{\mathbf{w}}_M^{(r)}\}$ ($\hat{\mathbf{w}}_m^{(r)} \in \{\mathbf{w}_1, \dots, \mathbf{w}_I\}$) を選択する。ここで M は楽譜中のコード変化の回数 + 1、 I はコードの種類数であり、 $\{C, C\sharp, \dots, B\} \times \{\text{major}, \text{minor}\}$ の 24 種類とした。また、 \mathbf{w}_i は各コードラベルに対応する基底スペクトルである。コードはセミマルコフモデル (SMM) に従う。

$$p(z_1) = \pi_{z_1}^{(r)}, p(z_m | z_{m-1}) = \psi_{z_{m-1}, z_m}^{(r)} \quad (1)$$

ここで $\pi_{z_1}^{(r)}$ と $\psi_{z_{m-1}, z_m}^{(r)}$ はそれぞれコードの初期確率とコード z_{m-1} からコード z_m への遷移確率を表す。また、コードのオンセット位置の遷移確率 $\phi_{u_{m-1}, u_m}^{(r)}$ は以下のように定式化される。

$$p(u_m | u_{m-1}) = \phi_{u_{m-1}, u_m}^{(r)} \quad (2)$$

ここで、 $u_m^{(r)} \in \{0, \dots, 15\}$ はコード z_m のオンセットビート位置を表す。

リードギターモデル リードギターの音高系列の生成過程を定式化する。リードギターの HSMM は、コード遷移モデルにより決定されたコード列の下でメロディ音符列 $\mathbf{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$ ($\mathbf{y}_n \in \{1, \dots, J\}$) を生成し、それに従って振幅スペクトル基底 $\hat{\mathbf{w}}^{(l)} = \{\hat{\mathbf{w}}_1^{(l)}, \dots, \hat{\mathbf{w}}_N^{(l)}\}$ ($\hat{\mathbf{w}}_n^{(l)} \in \{\mathbf{w}_1, \dots, \mathbf{w}_J\}$) を選択する。ここで N は

Estimation of Lead and Rhythm Guitar Scores for Music Audio Signals Based on a Factorial HSMM: Kentaro Shibata, Yoshiaki Bando, Yuta Ojima, Ryo Nishikimi, Katsutoshi Itoyama, and Kazuyoshi Yoshii (Kyoto Univ.)

楽譜中のソロギターの音符の個数, J はソロギターの音符の種類数であり, $\{E2, F2, \dots, C6\}$ の 44 種類とした. また, w_j は E2 から C6 の音高に対応する基底スペクトルである. リードギターの HSMM は階層 HSMM となっており, 上階はコード系列 Z を潜在変数系列としメロディの音符系列 Y を生成する HSMM, 下階はメロディの音符系列 Y を潜在変数系列として振幅スペクトル基底系列 $\hat{w}^{(l)}$ を選択する SMM となっている. メロディ列は SMM に従う.

$$p(y_1|z_1) = \pi_{z_{m(y_1)}, y_1}^{(l)} \quad (3)$$

$$p(y_n|y_{n-1}, z_{m(y_n)}) = \psi_{z_{m(y_n)}, y_{n-1}, y_n}^{(l)} \quad (4)$$

ここで $m(y_n)$ は y_n のオンセットが属する小節を表し, $\pi_{z_{y_1}}^{(l)}$ と $\psi_{z_{m(y_n)}, y_{n-1}, y_n}^{(l)}$ はそれぞれコードが与えられた元でのリードギターの初期確率と音高 y_{m-1} からコード y_m への遷移確率を表す. リズムギター同様に音符のオンセット位置の遷移確率 $\phi_{v_{n-1}, v_n}^{(l)}$ は,

$$p(v_n|v_{n-1}) = \phi_{v_{n-1}, v_n}^{(l)} \quad (5)$$

で与えられる. ここで, $v_n \in \{0, \dots, 15\}$ は音高 y_n のオンセットビート位置を表す.

音響モデル FHSMM によってリズムギターの基底列 $\hat{W}^{(r)}$, リードギターのメロディ基底列 $\hat{W}^{(l)}$ が選択された後に振幅スペクトログラムが生成される過程を定式化する. 音響モデルでは, 選ばれた基底それぞれにアクティベーション $h^{(r)}, h^{(l)}$ を掛け合わせたものの和を出力し, 以下の様に表される.

$$p(x_{ft}) = p(x_{ft}; \hat{w}_{f, n(t)}^{(l)} h_t^{(l)} + \hat{w}_{f, m(t)}^{(r)} h_t^{(r)}) \quad (6)$$

ここで $n(t), m(t)$ はそれぞれ t 番目の時間フレームが属する音符, コードのインデックスである.

2.2 推論

入力された振幅スペクトログラム X に基づき潜在変数系列 Y, Z を最尤推定する. 各遷移確率 $\psi^{(l)}, \psi^{(r)}, \phi^{(l)}, \phi^{(r)}$, 初期確率 $\pi_{z_{m(y_1)}, y_1}^{(l)}$ および基底 $W^{(l)}, W^{(r)}$ は事前学習する. Y, Z はビタビ探索とアクティベーション $h^{(l)}, h^{(r)}$ の最尤推定を交互に更新する. 出力楽譜は得られたコード・メロディ音高系列にアクティベーションの重みを掛け合わせたものに 16 分音符単位ピーク検出を行うことで推定する. ピーク箇所を音符のオンセットとし, オフセットを次の音符のオンセットとしてリズム・リードギター譜を得る.

3. 評価実験

提案法による楽譜推定精度を評価する. 楽曲は RWC データベース [4] のポピュラー音楽 100 曲のうち, 1 から 50 番の冒頭 30 秒を用いる. ビート時刻を得るために Florian ら [5] によるダウンビートトラッキングを行った. 遷移確率および初期確率は Mr.Children 9 曲, Superfly 15 曲, The Beatles 27 曲計 51 曲の楽譜を用いて事前学習した. リードの単音は一般的なギターの音域である E2 から C6, コードはギターの一一般的なローコードで最低音が必ずルート音になるものとして, MIDI ギター音源より基底を事前学習した.

提案法を用いて推定された楽譜のコンデンススコアとデータベース内の XML[6] より作成したコンデンススコアの音高の一致率を F 値により比較した. ただし, ボーカルとドラムパートは除いた. 結果を表 1 に示す. また,

表 1: 音高推定精度 [%]

	F-値	適合率	再現率
一致率	42.92	45.53	41.76
構成音一致率	71.91	77.12	68.99

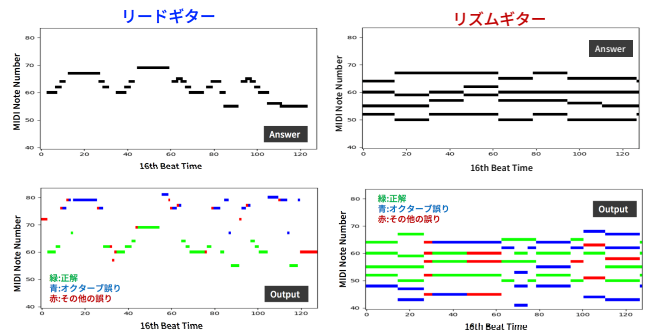


図 2: パート譜推定結果

RWC の 12 番の MIDI データからリード・リズムパートを担う 2 つのギターパートを抜き出し, 音響信号を再合成したものを対象に, 本手法を用いて楽譜を推定した. 比較結果を図 2 に示す.

図 2 に示すように, メロディパートとリズムパートを分けて推定可能なことを確認した. 一方, メロディパートは追えているものの, オクターブ誤りをしているため, 一致率が悪化しているものが見られた. 実際は楽器が鳴っていない区間でも, 本手法ではコードとメロディを出力し続けるため, 一致率が低いと考えられる. また, 実際はパワーコードやハイコード等で演奏している場合も本手法ではローコードとして採譜してしまことでオクターブ誤りが生じている.

4. おわりに

本稿では, 奏法の違いに基づき FHSMM を用いて 2 本のギターパートを推定する手法について提案した. ギターのみでの再構成音ではメロディとコードを分離して推定可能であることを確認した. 一方, コードの構成音が固定されていることによるオクターブ誤りや, 無音状態でもコード・メロディを出力する挿入誤りが見られた. 今後は, 同一コードの異なるフォームを正しく認識するために, 基底のを適応的に更新することや, 音を出力しない休符の状態を扱えるモデルに拡張することで, 推定結果の精度向上を図る予定である. また, リードギターと同様にベースギターも扱い, 統一的な枠組みでリード・リズム・ベースギターを採譜するモデルへの拡張を目指す. 謝辞 本研究の一部は, JSPS 科研費 26700020, 16H01744 および JST ACCEL No. JPMJAC1602 の支援を受けた.

参考文献

- [1] S. Ariga *et al.*: "Song2Guitar: A Difficulty-Aware Arrangement System for Generating Guitar Solo Covers from Polyphonic Audio of Popular Music," *ISMIR*, p568-574, 2017.
- [2] A. Jansson *et al.*: "Singing Voice Separation with Deep U-Net Convolutional Networks," *ISMIR*, 323-332, 2017.
- [3] N. Ono *et al.*: "Harmonic and Percussive Sound Separation and Its Application to MIR-Related Tasks," *Advances in Music Information Retrieval*, 213-236, 2010.
- [4] M. Goto *et al.*: "RWC Music Database: Popular, Classical and Jazz Music Databases," *ISMIR*, 287-288, 2002.
- [5] K. Florian *et al.*: "Downbeat Tracking Using Beat-Synchronous Features and Recurrent Networks," *ISMIR*, 129-135, 2016.
- [6] M. Goto *et al.*: "AIST Annotation for the RWC Music Database," *ISMIR*, 359-360, 2006.