

調波音・打楽器音分離手法を用いたギター・ベースギターの自動採譜*

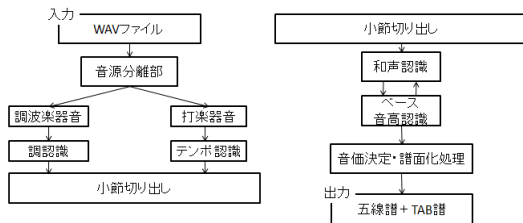
木村拓也 (法政大学 情報科学部), 伊藤 克亘 (法政大学 情報科学部)

1 まえがき

本稿では楽器初心者の練習の補助のために、音楽音響信号から楽器初心者でも理解しやすい五線譜及びTAB譜の出力を行うシステムを提案する。研究対象とする楽曲は、楽曲全体を通してテンポが一定であり、変拍子や転調を含まないものとする。

2 システム

概要 下図でシステムフローを示す。音源分離部では、FitzGerald[1]の手法のサンプリングレートを16000Hz、FFT sizeを512点、pop sizeを128点に変更し、調波・打楽器成分の分離を行った。



高頻出周波数の抽出 楽曲は固有の調に基づく音階が使用される。そこで、楽曲全体の音響信号を分割し、各フレームでスペクトルを求める。そのスペクトルから閾値以上のパワーの周波数とそのパワーを検出し、全ての音名の使用率を求め、上位7音と各調の音階7音との相関が一番高くなる調を楽曲の調と推定する。

テンポ認識・小節切り出し 譜面化には音価と小節の区切りを決定するために曲のテンポを認識することが必要不可欠である。そのため、分離した打楽器成分のパワーのピークに対し自己相関を用いて音量変化の周期を検出することによって、ビート間隔の候補を推定しテンポを決定する。テンポの認識は楽器初心者でも比較的簡単に行えるため、検出されたテンポが誤っていると判断した場合には候補の中から正しいテンポを選択し結果を修正できる。

また、小節の開始点を探すために、認識したテンポからビート間隔でフィルタを作成し、そのフィルタを1点ずつずらしながらパワーとの内積を求めた。開始点の候補を複数検出し、実際にユーザーに切り出した音声を聞かせ、開始点を選択させる。

音高認識モジュール

問題の定式化 n 個の小節に分割した信号に対して、各小節で1つの和音と、小節をさらに m 個に分割した各時刻でのベース音高を認識する。その際のコード系列を $C = [c_1, c_2, \dots, c_n]$ とし、各 C_n に対するベース音高

系列を $B_n = [b_1, b_2, \dots, b_m]$ とする。 B_n を求める際には、各フレームでスペクトル特徴 s と、事前学習させた各コードで使用されるベース音高分布 L の2つを用いる。これらから、ベース音高系列に関する事後確率 $P(B_n|s, L)$ を考え、与えられた s 及び L に対する最大事後確率 (MAP) 推定問題としてベース音高系列認識を定義する。すなわち、 $\arg\max_B P(B|s, L)$ が求めるべきベース音高系列である。

各コードで使用されるベース音高分布の学習 各コードで使用されるベースの音高は、音楽理論的観点からある程度定めることができる。よって、各コードで使用されるベース音高分布を事前学習で求め、ベース音高系列認識に利用する。まず、コードが一つのみ使われる解析区間を設定し、各解析区間でのコードの正解データを既存の楽譜から作成した。ベース音高については、TAB 譜共有サイト UltimateGuitar で信頼度の高いもののベースパートの MIDI データや、既存の楽譜から手入力した MIDI データを利用した。学習させる際には、調によって各コードの位置付けが変わってしまう点を解消するために、各調のトニック音を基準にし、そこから何音ずれるかを計算する。そして、各コードの区間でベースの音高と音価を計算し、音名を表す12次元の配列に音価を足していくことで、12次元のヒストグラムからなる分布を作成する。学習させたベース音高分布 L はメジャー、マイナーの24種である。学習に用いた楽曲は、洋・邦楽の rock, pops のジャンルから抜粋した14曲である。

倍音フィルタを用いたベースライン推定 ベースラインを推定する場合、自己相関や通常のFFTのスペクトルから求める方法では、周波数帯域が重なりやすいバスドラムなどの影響を大きく受けてしまう。そこで、下図のように、ベースで 사용되는基本周波数とその7次までの倍音からなるフィルタ HF を $C0$ から $B2$ までの36種生成する。



ベース音高候補系列 B は倍音フィルタ HF を周波数スペクトル s に適用した際に、最も値が大きくなるものを探せば良いので、 $P(s|HF)$ が最大になる HF を求める。ここで、 $P(s|HF)$ は多項分布だとすると、その確率は式1で表される。ここで、 ω は各周波数ビンを表し、倍音の次数は実験結果から一番精度の良かったものを使用した。

$$P(s|HF) = P(s|B) = \prod_{\omega} (HF_{\omega})^{s_{\omega}} \quad (1)$$

* Automatic Transcription of Guitar/Bass Guitar Part Using Harmonic/Percussive Sound Separation Method.: Takuya Kimura (Hosei Univ.) et al.

コード認識 コード認識を行うために、ベース音高とクロマベクトルを用いた手法を提案する。音階名で表される基本情報(例えば「C3」)はオクターブ(「3」)と半音階名(「C」)を指定することで一意に定めることができる。このうちオクターブ違いの同じ音階の成分を全て足しあわせて、1オクターブ内の半音階の12音の強さを表したものがクロマベクトルである。

2.5節で切り出した区間からスペクトログラムを求める。ここで、各時間ビンに対し50セント幅で周波数をクロマベクトルに変換し、そこからパワーの強い3音を求めることで、コードの候補を推定する。さらに、ベース音高は一般的にそのコードの根音であることが多く、学習データからもそれがわかるので、求めたベースの根音と楽曲の調から考えられるコードを、クロマベクトルに1/区間数の重みを足すことによってコードを推定した。

最大事後確率の計算 これらから、 $\arg\max_B P(B|s, L)$ を計算する。

$$P(B|s, L) = P(B|s)P(B|L) \quad (2)$$

ここで、 $P(B|s)$ についてベイズの定理から次式が成り立つ。

$$P(B|s) = \frac{P(s|B)P(B)}{P(s)} \quad (3)$$

この時、 $P(s)$ は定数、 $P(B)$ は一様分布で、ベース音高系列 B は倍音フィルタ HF と等価であるので、次式が成り立つ。

$$P(B|s) \propto P(s|HF(B)) \quad (4)$$

これらを用いて、式2の右辺の対数を取り B を求める。

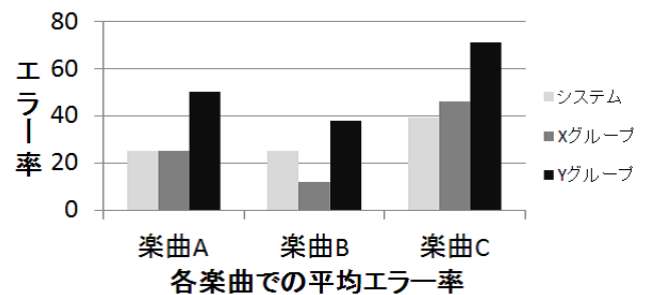
$$\arg\max_B \left(\sum_{\omega} s_{\omega} \log HF(B) + \log P(B|L) \right) \quad (5)$$

楽譜生成 楽譜生成のために、ノートナンバー、発音時刻、消音時刻を指定する必要がある。各区間で認識したコードは、指定した音価で出力する。ベース音については、フレームごとの解析結果を、指定した音価分の長さでモードを取り、音高を一意に定める。

3 評価実験

認識率評価 3つの楽曲に対し、評価実験を行った。ベースの音高については、楽譜のノートナンバーのズレを誤検出とする。また、本稿で取り扱うコードは、メジャー、マイナーの24種類とし、その他のコードはこれらのコードの派生コードとし、sus4やaugはメジャーに、m7やdimはマイナーに分類する。また、認識率は16分音符ごとの正解率とする。ベース音、コードの平均正解率はそれぞれ68%、84%となった。精度が悪かった楽曲は、ベース音はコードのルート音を弾くという制約から逸脱したフレーズであったため、コードの誤認識を起こし、最大事後確率の計算にも悪影響を及ぼした。今後は、コード認識モジュールでは別の制約を定める必要がある。

有効性評価 本システムが楽器初心者の採譜支援に有効であるか確認するために、4人の被験者を用いた評価実験を行った。評価には認識率調査を行った3曲を用い、各楽曲で制限時間を15分に定め、本システムを使用するXグループと使用しないYグループ、それぞれ2人ずつの間での平均エラー率を比較した結果を下図に示す。ただし、ここではコード認識については、派生コードをメジャー・マイナーとして認識することは不正解とし、楽譜生成の際の音価は8分音符に指定した。結果から、X,Yそれぞれのグループの平均エラー率は29%、53%となり、本システムを利用することによってエラー率が42%減少することが確認された。また、被験者の採譜結果やコメントから、ベースについては、主にルート音を弾いている楽曲は7割程度の正解率であればシステムの出力譜面の誤認識部分を判断しやすいため、スムーズな採譜が行えた。しかし、音高が細かく変化するフレーズでは、誤認識部分の判断が難しく、効果的な支援には繋がらなかった。ギターコードに関しては、音色が大きく歪んでいると出力譜面との聴き比べが困難になる為、正解を出力していてもそれを誤認識と判断するケースがあった。これらは解析結果の信頼度が表示されないことが原因に挙げられ、今後は信頼度を表示することによって、より効果的な採譜支援が行えることを示唆している。



4 あとがき

本研究では、和声認識やテンポ検出、音高推定の性能向上のために調波音・打楽器音分離を主とした音源分離を行った。また、ベース音高とクロマベクトルを用いたコード認識、事前学習させた各コードのベース音高分布と倍音フィルタによるベース音の基本周波数認識を行った。これらの情報から五線譜とTAB譜を生成した。3曲の楽曲に対し自動採譜を行ったところ、平均76%の認識率を得た。また、本システムを利用しない被験者2人と利用した被験者2人で採譜を行ったところ、本システムを利用した際にはエラー率が42%減少し、本システムが楽器初心者の採譜支援に有効であることを確認した。

参考文献

- [1] Derry FitsGerald, "HARMONIC/PERCUSSIVE SEPARATION USING MEDIAN FILTERING", DAFX10, Graz, Austria, September 6-10, 2010.