Vol.2020-MUS-129 No.2
2020/11/2

演奏タイミングのズレを含む混合音に対する マルチラベルビートトラッキング

福谷和貴 [1]　酒向慎司 [1]

概要：本研究では，ビートラベル列が楽器ごとに存在し，マルチラベルとなっている演奏データに対して，それぞれのビートラベル列を認識するマルチラベルビートトラッキングを検討する．楽器演奏の初心者は一定のリズムで演奏しているつもりでも，タイミングがズレてしまうことがある．特に複数の楽器で合奏している場合，自分の演奏がズレているのか判断するのは困難である．そこで，演奏のタイミングがずれてしまう初心者に対して，演奏タイミングを示すことによって，練習支援を行うことができると考える．手法としては，各楽器に対応したビートラベル列を持つ複数楽器の音が混合された音響信号に対して，マルチラベルでのビートトラッキング技術を用いることで楽器ごとの演奏タイミングを認識する．まず，初心者の演奏を模擬した演奏データを作成する．次に，LSTM を用いたビートトラッキング手法をマルチラベルに拡張する．評価実験の結果，従来手法を組み合わせた方法でギター，ドラムについてそれぞれ F 値が 0.578，0.760 であったのに対し，マルチラベルビートトラッキングを行うことでそれぞれ F 値が 0.603，0.754 となった．

キーワード：ビートトラッキング，LSTM，練習支援，ギター演奏

1. はじめに

楽器を演奏する際，一定のリズムで演奏するには一定の技量が必要であり，初心者にとっては困難なことである．また，何種類かの楽器で合奏をしていると，他の楽器の音で自分の音がわからなくなってしまうため，一定のリズムで演奏できているか判断するのはより困難なこととなる．そこで，合奏において楽器ごとの演奏タイミングを示すことで，練習支援が可能であると考えられる．

演奏タイミングに着目した練習支援の研究は従来から行われている [1]．このような研究では練習支援の対象として MIDI ドラムなどの電子楽器が用いられており，その演奏情報と楽譜とを比較することで練習支援を行っている．電子楽器を用いることで特定の楽器の演奏タイミングを簡便に取得することができるが，楽器の種類が限定され，任意の楽器に対応した演奏支援システムは難しい．

そこで本研究では音響信号からのビートトラッキングを利用した練習支援に着目する．合奏において楽器ごとのビートラベル列を認識することで，楽器ごとの演奏タイミングを示すことを目標とする．しかし，ビートトラッキングの一般的な問題設定は，ビートラベル列が 1 つである．

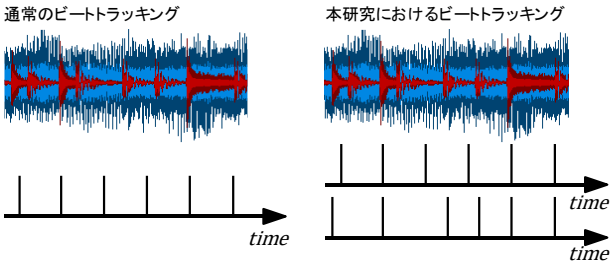


図 1 通常のビートトラッキングと本研究との違い

図 1 に通常のビートトラッキングと本研究との違いを示す．筆者の知る限り音響信号から楽器ごとのビートラベル列を得る様なマルチラベルのビートトラッキングは存在しない．従って，問題設定の簡単化のため，本研究の問題設定では対象をギターとドラムの 2 つの楽器での演奏に限定する．また，ギターの演奏のみが初心者（一定のリズムで演奏できない人）の演奏であるとする．

本研究では初心者を模擬した演奏データを作成して，楽器ごとのビートラベル列を生成することでデータセットの構築を行う．次にビートトラッキングの従来手法をマルチラベルに拡張し，評価実験を行う．

本稿では 2 節で一般的なビートトラッキングについて記述する．3 節で初心者の演奏を模擬したデータセットの構築について述べた後，4 節で既存手法の拡張方法について説明する．評価実験について 5 節で述べ，6 節にてまとめを行う．

2. ビートトラッキング

ビートトラッキングとは人間が音楽を聴きながら手拍子をするような位置（ビート位置）を音楽音響信号から推定する技術である．音楽情報処理の基礎技術の一つであり，音楽情報処理の国際コンペティションである MIREX（the Music Information Retrieval Evaluation eXchange）[2]において Audio Beat Tracking として，主要なタスクの一つである．

ビートトラッキングの基本的な手法を説明する．まず，音楽音響信号から，音の立ち上がりを表す特徴量を抽出す

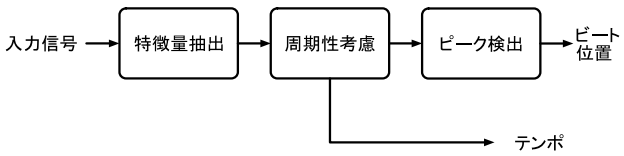


図 2 基本的なビートトラッキング手法

1 名古屋工業大学大学院
Nagoya Institute of Technology

ⓒ2020 Information Processing Society of Japan

表 1 データセット構築において除外した曲一覧

楽曲番号	曲名
No. 71	月のように
No. 72	Heart to Hurt
No. 73	Miss Maria
No. 74	きみの色
No. 75	遠い街へ
No. 76	誓い
No. 77	愛してる
No. 78	雲
No. 79	Together
No. 80	さがしもの
No. 99	Once and For All

る．抽出した特徴量から周期性を考慮することで楽曲のテンポを推定し，抽出した特徴量のピーク位置と推定したテンポを考慮しビート位置が推定される．図 2 に基本的なビートトラッキング手法の概要図を示す．

近年はこのような信号処理的な手法から機械学習を用いた手法が主流になってきている．特に，時系列を考慮できる RNN を用いた研究が多くなされている．MIREX においては，RNN の一種であり，長期的な依存関係を学習することのできる LSTM を用いた手法が好成績を収めている．Böck らは 3 層の Bidirectional LSTM (BLSTM) を用いたビートトラッキング [3][4][5]を提案している．この研究では，BLSTM を用いて楽曲のフレームごとのビートである確率を示す beat activation 関数を算出し，この関数と自己相関関数 [2]や Dynamic Bayesian Network [3][4]を使用することでビート位置を推定している．

3. データセットの構築

本研究では，初心者を模擬したデータセットの作成を行った．初心者の演奏を模擬するために，既存のデータセットの演奏データ（MIDI ファイル）を操作した．データセットとして RWC 研究用音楽データベース（RWC-MDB） [6]のポピュラー楽曲 100 曲を使用した．

本研究においては対象楽器をギターとドラムのみに限定しているため，MIDI ファイルからギターとドラムのトラックのみ抽出した．この際，ドラムトラックがないものについて今回はデータセットから除外することとし，その結果，データセットの曲数は 89 曲となった．表 1 に今回除外した曲一覧を示す．

3.1 無音区間の扱い

様々な楽器で演奏されている MIDI ファイルからギターとドラムのトラックを抽出した結果，ギターのソロパートなど，一方の楽器のみ演奏しており，もう一方の楽器の演奏データでは無音の区間が生じる可能性がある．このような無音区間は通常のビートトラッキングでは生じないため，扱いを検討する必要がある．本研究では BLSTM における学習での悪影響を考慮し，それぞれの楽器に対して，無音区間ではビートが存在しないビートラベル列を作成した．

3.2 タイミングのズレの付与

どのような操作を行えば初心者を模擬することができるのかについても検討する必要がある．本研究では，ギタートラックに対して初心者の演奏を模擬するために，ギタートラックのイベントのタイミングをランダムにずらす以下の操作を行った．

ギタートラックの n 番目のイベントを E_n とする
E_n のズレを以下の 3 種類から同確率で決定
[前にズレる・ズレない・後ろにズレる]

- 前にズレるとき
  E_{n-1} の t 秒後から E_n の t 秒前の間でランダムにずらす
- 後ろにズレるとき
  E_n の t 秒後から E_{n+1} の t 秒前の間でランダムにずらす

ただし，イベントの順番は変わらないものとする．70 ms 以内の間に生じた 2 つの音は 1 つに聞こえる [7]ため，本研究では $t = 0.07$ とする

こうして作成された MIDI ファイルから WAV ファイルを作成することでデータセットを構築し，総時間としては 20,791 秒となった．

4. 提案手法

本研究においてはBLSTMを用いた先行研究 [3][4]を参考にマルチラベルビートトラッキング手法を提案する．

4.1 特徴量抽出

BLSTM の入力として異なる窓幅で算出したメルスペクトログラムを連結したものを用いる．図 3 に入力する特徴量の一例を示す．音楽音響信号のサンプリング周波数を 44,100 Hz とする．特徴量を算出する際の窓幅を 4,096, 2,048, 1,024 サンプル（92.8，46.4，23.2 ms）とし，シフト幅を 441 サンプル（10 ms）とする．入力次元削減のため，周波数帯域を[30, 17,000 Hz]に限定する．4,096，2,028，1,024 サンプルの窓幅で算出された特徴量において，1 オクターブあたり 12，6，3 バンドのフィルターを使用する（図 3 の(a)，(c)，(e)に対応）．また，動的特徴量としてメルスペクトログラムの 1 次差を特徴量に連結する（図 3 の(b), (d), (f) に対応）．結果として 381 次元の入力特徴量が算出される．

ⓒ2020 Information Processing Society of Japan

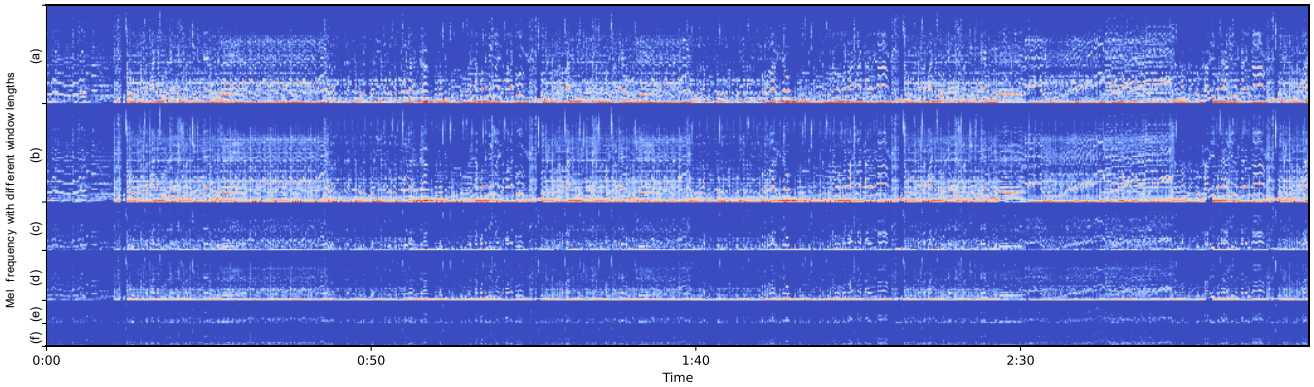


図 3 入力特徴量の一例

4.2 BLSTM の詳細

マルチラベルに拡張する前の BLSTM について説明する．本研究では 3 層の BLSTM を用いる．隠れ層は 1 層につき 25 の LSTM ユニットを持つ．入力層が 381 次元であり，出力層は“beat”と“non-beat”の 2 次元の層である．活性化関数として softmax を用い，BLSTM の出力はあるフレームにおけるビートである確率となる．これを時系列に並べたもの（beat activation 関数）と [8]で提案されている HMM を用いることで音響音楽信号におけるビート位置を推定した．本研究では madmom [9]を使って HMM を実装した．

4.3 BLSTM のマルチラベルへの拡張

前節で述べた BLSTM の出力層をギターについて，ドラムについての 2 層用意することで，マルチラベルに拡張する．結果として出力は 4 ユニットになる．本研究の提案手法の概要図を図 4 に示す．

4.4 BLSTM の学習

データセットに対して 5 倍の交差検証を行った．学習データの 2 割を検証データとして使用し，残りのデータで学習を行う．ネットワークの隠れ状態を[−0.1，0.1]の範囲でランダムに初期化を行う．最適化関数として Adam を用い，学習率を 0.01 とする．学習において，20 エポックの間に検証データに対する損失の改善が見られない場合，学習を終了する．その後，学習率を 0.001 に変更し，同じ終了条件において再学習を行う．

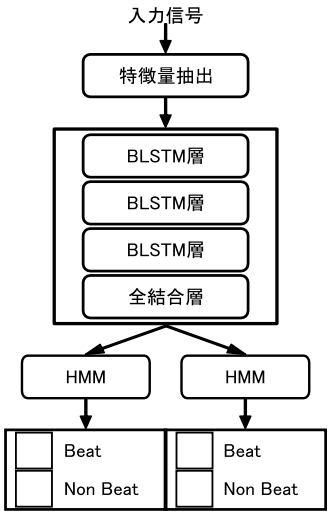


図 4 提案手法の概要図

5. 評価実験

他のビートトラッキングの研究と同様に，本研究では特に記述がない場合，評価実験において±70 ms の誤差基準 [10]を使用する．提案手法の評価を行うため，2 つの評価実験を行った．

5.1 実験 1

他のビートトラッキング研究と異なり，本研究では演奏楽器をギターとドラムに限定している．これは曲の構成要素が限定されているためビートトラッキングの精度への何らかの影響があると考えられる．従って，3 節で構築したデータセットではなく RWC-MDB のポピュラー楽曲 100 曲からギターとドラムを抽出したデータ 89 曲に対してビートトラッキングを行い，その影響の有無を確認する．この時，除外した 11 曲はデータセットを構築した際に除外した 11 曲と同じである．

4.1 節，4.2 節，4.4 節で述べた条件で実験を行った結果，F 値が 0.760 であった．すべての楽器の演奏データに対して同じ実験条件で実験を行った結果，F 値が 0.772 であったため，演奏楽器を限定したことによるビートトラッキングの精度への影響は無いことが確認された．

5.2 実験 2

次に，3 節で構築したデータセットに対して提案手法の有用性を確認する．

5.2.1 実験条件

比較手法として，本研究の問題設定に対して，従来手法を組み合わせる手法を用いる．具体的には，混合音に対し

ⓒ2020 Information Processing Society of Japan

表 2 誤差基準を変化させた実験結果（F 値)

	誤差基準 ：35 ms		70 ms	140 ms
提案手法	ギター	**0.577**	**0.603**	**0.696**
	ドラム	**0.744**	**0.754**	**0.772**
比較手法 1	ギター	0.540	0.578	0.704
	ドラム	0.753	0.760	0.764
比較手法 2	ギター	0.346	0.404	0.595
	ドラム	0.758	0.764	0.769

て Harmonic-percussive source separation （HPSS）[11]を行い，調波音と打楽器音を得る．得られた音それぞれに対して，4.1 節，4.2 節，4.4 節で述べた条件で学習，推定を行う（比較手法 1)．しかし，HPSS での音源分離は完璧ではなく，完全な音源分離は不可能である．従って更なる比較手法として，理想的な音源分離の場合の手法（比較手法 2）においても実験を行う．比較手法 2 においては，3 節で構築したデータセットを使用せず，ギタートラックのみ，ドラムトラックのみの MIDI ファイルから作成したギター，ドラムのソロ音源を使用する．3，4 節で述べた通り，89 曲（20,791 秒）のデータを使用し，5 倍の交差検証を行った．

5.2.2 実験結果

実験結果を表 2 に示す. 楽器ごとに着目して考察を行う．ドラムに着目すると，提案手法，比較手法に関して大きな差は見られず，実験 1 の結果と比較しても，遜色ない結果が出ている．

次にギターに着目する．比較手法 2 においてギター音の精度が低いのは，ギター演奏がドラム演奏と比べてオンセット数が少なく，得られる情報が少ないため学習が上手く行えていないためであると考えられる．比較手法 1 においては HPSS における音源分離は完璧ではなく，ギター音にドラム音が混入しているため，比較手法 2 より高い精度が出ているのだと考えられる．提案手法において最も高い精度が出ており，ギターのソロ演奏では上手く学習できないことでも，入力に混合音を使用し，ドラムとギターのビートラベルを用いてマルチラベル学習を行うことで，正しく学習ができることが示され，提案手法の有効性が示された．

ただし，提案手法においてもギターについて高い精度が出ているとは言えず，出力の分析を行ったところ，全体的な傾向として，ある程度正しいビート位置の推定は行えているものの，ズレに対して対応できていないという傾向があった．ズレに対する頑健性を高め，初心者の演奏タイミングの不一致を正しく検知することが今後の課題である．

6. まとめ

本研究では，演奏タイミングのズレを含む混合音に対するマルチラベルビートトラッキングを提案した．マルチラベルビートトラッキングを行っている研究が筆者の知る限り存在しないため，簡単化のため問題設定を，ギター，ドラムの 2 種類の楽器に限定し，初心者の演奏をギターのみとした．MIDI ファイルを用いて初心者を模擬した演奏データを作成し，データセットの構築を行った．また，BLSTM を用いたビートトラッキング手法をマルチラベルに拡張した．評価実験を行い，従来手法を組み合わせた手法と比較を行った結果，マルチラベルでビートトラッキングを行うことの有効性は示されたものの，十分な精度であるとは言えない結果となった．

今後の方針としては，学習データに対して，ピッチシフトやタイムストレッチなどの一般的なデータ拡張や，疑似初心者データを作成する際のパラメータを変更するなどのデータ拡張を行い，精度を高めていく方針である．また，疑似的なデータでの評価実験だけでなく，ギターの演奏データ収録を行い，実際の初心者の演奏データでの評価実験も行う予定である．

謝辞 本研究のビートトラッキングの実装には AGH 科学技術大学（ポーランド）の Jakub Gałka 博士の助言を受けた．

参考文献

[1] 岩見直樹 ，三浦雅展，“MIDI 楽器を用いたドラム演奏練習支援システムの提案,” 情報処理学会研究報告音楽情報科学 (MUS), pp. 85--90, 2007.
[2] MIREX, https://www.music-ir.org/mirex/wiki/MIREX_HOME,（参照日：2020/10/06）
[3] S. Böck , M. Schedl, “Enhanced beat tracking with context-aware neural networks,” Proc. Int. Conf. Digital Audio Effects, pp. 135--139, 2011.
[4] S. Böck, F. Krebs , G. Widmer, “A Multi-model Approach to Beat Tracking Considering Heterogeneous Music Styles,” ISMIR, pp. 603--608, 2014.
[5] S. Böck, F. Krebs , G. Widmer, “Joint Beat and Downbeat Tracking with Recurrent Neural Networks,” ISMIR, 2016.
[6] M. Goto, H. Hashiguchi, T. Nishimura and R. Oka, "RWC Music Database: Popular, Classical, and Jazz Music Databases," ISMIR, vol. 2, pp. 287--288, 2002.
[7] S. Dixon, "Automatic extraction of tempo and beat from expressive performances," Journal of New Music Research, vol. 30, no. 1, pp. 39--58, 2001.
[8] F. Krebs, S. Böck , G. Widmer, “An Efficient State-Space Model for Joint Tempo and Meter Tracking,” ISMIR, pp. 72--78, 2015.
[9] S. Böck, F. Korzeniowski, J. Schlüter, F. Krebs , G. Widmer, “Madmom: A new python audio and music signal processing library,” Proceedings of the 24th ACM international conference on Multimedia, pp. 1174--1178, 2016.
[10] M. E. P. DAVIES, N. DEGARA and M. D. PLUMBLEY, "Evaluation methods for musical audio beat tracking algorithms," Queen Mary University of London, Centre for Digital Music, Tech. Rep. C4DM-TR-09-06, 2009.
[11] Derry Fitzgerald. Harmonic/percussive separation using median filtering. International Conference on Digital Audio Effects (DAFX), pp. 1--4, 2010.

ⓒ2020 Information Processing Society of Japan