

# ライブ演奏楽曲の楽曲同定におけるビート情報取得手法の検討及び改善 An Examination and Improvement of a Way to Get Beat Information in Audio Live Version Identification

石倉 和将†  
Kazumasa Ishikura

甲藤 二郎†  
Jiro Katto

## 1. まえがき

近年 iPod などの普及に伴い、ユーザーが扱う楽曲数が格段に増加した。一方でユーザーが自身の嗜好に合った楽曲を見つけることが困難になり、この問題を解決する手段として楽曲推薦や楽曲検索、そして楽曲同定の研究がある。

世に出回っている楽曲同定システムとして Shazam[1]や Sound Hound[2]といったものが挙げられるが、これらはカバーソングやライブバージョンといった「原曲に似ているが同一ではない楽曲」には対応できていない。この問題に対し、筆者らはライブバージョンの楽曲、特に実環境におけるライブ演奏楽曲に着目し、それらを対象とした楽曲同定システムを提案した[3]。

しかし提案した手法ではビート情報の取得が正しく行われなかった場合に誤同定が発生する可能性が高いという問題があった。そのため本研究ではビート情報の取得手法の違いによる同定精度の変化を考察し、ライブバージョンの楽曲同定に最適なビート情報取得手法及び楽曲同定手法を提案する。

なお本論では、スタジオレコーディングの CD 音源を原曲、原曲アーティスト以外の奏者による演奏・歌唱音源をカバーソング、原曲アーティストが生演奏し、マスタリングやミキシング等の処理が加えられた音源をライブバージョン、そのような処理を加えられずライブ会場において観客が耳にする音源を実環境ライブバージョンと呼ぶ。

## 2. 従来手法

ライブバージョンに対する楽曲同定の研究は MIREX[4]において Audio Cover Song Identification のタスクに内包されている。カバーソングの楽曲同定の研究において主流となっている手法は 2006 年に Ellis が提案した、ビート同期をしたクロマベクトル(ビート-クロマ特徴量)を用いてマッチングを行うもの[5]である。Ellis の提案手法ではビート情報を取得する方法として、動的なビートトラッキング手法[6]が用いられている。

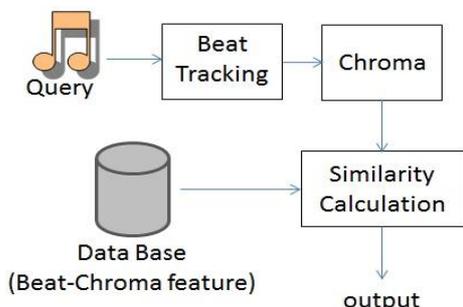


図 1: Ellis 手法[5]の処理の流れ

しかし後の調査により、楽曲同定の精度が動的なビートトラッキングの成否にかなり依存することが判明し、その後は†早稲田大学 基幹理工学研究所, Waseda University

動的なビートトラッキングへの依存性を小さくするような手法[7][8]が提案されるようになった。一方で、Riley らの論文[9]により、ライブバージョンとカバーソングには差異があり、同定精度にも隔たりがあることが知られている。

## 3. 予備実験

ビート情報取得手法の違いによる楽曲同定精度の違いを観察するため、2006 年に Ellis が提案したシステム[5]のビート情報取得部分を変更して同定実験を行った。

### 3.1 データセット

実験にはデータセットとして自身で収集した楽曲群(原曲 100 曲、カバーソング 50 曲、ライブバージョン 50 曲、実環境ライブバージョン 50 曲)を用いた。mp3 形式であり、サンプリングレートは 44.1kHz、ビットレートは 128kbps である。

### 3.2 ビート情報取得手法

今回用いたビート情報取得手法は Ellis の動的なビートトラッキング手法 2 種類[6][10]と、音響特徴量を用いたトラッキングを行わずあらかじめテンポを仮定してビートを生成する手法である。

### 3.3 実験結果&考察

図 2 に各ビート情報取得手法による楽曲同定精度を示す。この表において Bt は手法[6]、Bt2 は手法[10]、その他の数値は楽曲のテンポ(BPM)をその数値と仮定した場合を示している。

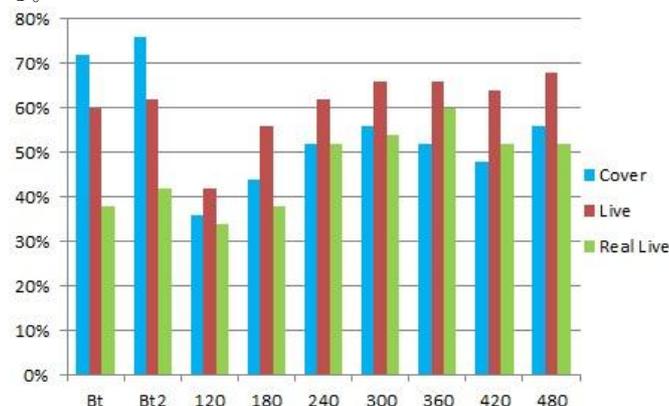


図 2: 各ビート情報取得手法による楽曲同定精度の変化

図 2 より、カバーソングにおいては動的なビートトラッキングを用いた場合の同定精度が高く、ライブバージョン及び実環境ライブバージョンにおいてはテンポを仮定した場合の同定精度が高いという結果となった。これは、カバーソングは原曲と違ったテンポで演奏される場合が多い一方で、ライブバージョンでは原曲と同様のテンポで演奏される場合が多いことや、実環境ライブバージョンについては実環境特有の残響や各楽器の音量バランス、観客の歓声・拍手といったものの影響で正しくビートトラッキングが為されないからだと考えられる。また、実際のポピュラー音楽のテンポは BPM60 から 240 あたりが多いのだが、この結果からすると本来のテンポより細かくビート-クロマ特徴量を取得した方が、同定精

度が高いようである。しかしあまり細かくしすぎると楽曲間の類似度が全体的に低くなってしまい、誤同定が発生してしまう。

#### 4. 提案手法

カバーソングには動的なビートトラッキング、ライブバージョンにはテンポ仮定が有効、という予備実験の結果を踏まえ、図 3 のような楽曲同定手法を提案する。

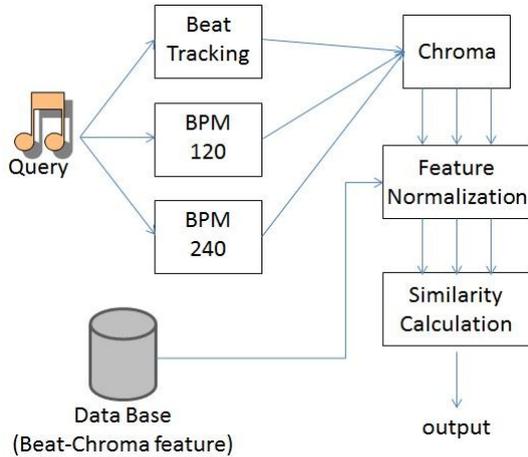


図 3: 提案手法の処理の流れ

このようにビート情報取得部分の処理でビートトラッキングとビート仮定の両方を行うことでライブバージョンの同定精度の向上を図る。動的なビートトラッキングを用いているためカバーソングの同定も可能である。並列化処理はビート情報を取得する部分から類似度計算まで行われ、得られた 3 つの同定結果から最も類似度のスコアが高い楽曲及び類似度のスコアを最終的な同定結果としている。類似度の計算には手法[5]の相互相関と巡回相互相関を組み合わせたものを利用しており、この計算方法は多くの従来手法で利用されている。

並列処理の組み合わせは多数存在するが、各組み合わせで同定精度を観察した結果、図 3 のように動的なビートトラッキング 1 つ(手法[10])と BPM を仮定する手法 2 つを組み合わせた場合が最もよい同定精度となった。また、2 つの BPM を仮定する手法はどのような組み合わせであっても同様の精度となったため、処理時間を考慮して仮定する BPM は 120 及び 240 が最適という結論に至った。図 2 によれば、仮定する BPM は 120 や 240 より高くすることで正しく同定される楽曲は増加するという事になっているが、その増加した楽曲はほとんどが動的なビートトラッキングによって正しく同定されている楽曲であるため、最終的な結果にあまり影響を与えない。

#### 5. 実験

前述したデータセットを用いて提案手法の楽曲同定精度を観察した。比較に用いる手法は Ellis が 2006 年に提案した手法[5]と、筆者自身が以前に提案した手法[3]の 2 つである。

表 1: 従来手法と提案手法の同定精度の比較

	手法[5]	手法[3]	提案手法
Cover	72%	70%	<b>92%</b>
Live	60%	72%	<b>88%</b>
Real-Live	38%	76%	<b>82%</b>

ビート情報の取得部分の処理を並列化することで表 1 のようにカバーソング、ライブバージョン、実環境ライブバージョンそれぞれに関して同定精度を向上することができた。特に実環境ライブバージョンの同定精度が著しく向上しており、動的なビートトラッキングで正しいビートを取得することが困難であることがわかる。並列化した処理のそれぞれの貢献度について、実環境ライブバージョンにおいては動的なビートトラッキングによる同定精度を基準とすると、BPM120 では新たに 22% の同定が成功し、BPM240 では 32%(うち BPM120 と重なる楽曲が 14%)の同定が成功し、それらを組み合わせることで最終的に 82% という精度が得られた。

#### 6. まとめ

本研究ではビート情報の取得手法の違いによる、それぞれの種類の楽曲の同定精度を評価し、いくつかのビート情報取得手法を組み合わせることで並列処理を行う楽曲同定手法を提案した。この結果、それぞれの種類の楽曲に対して従来手法より同定精度を向上することができた。特に大きな同定精度の向上が得られたのは実環境ライブバージョンであり、実環境特有の成分が従来手法には大きく影響を与えている一方で、提案手法ではテンポを仮定してビート-クロマ特徴量を取得する処理を加えることでビートトラッキングが失敗しても誤同定が発生しないようにすることができた。

たいていの楽曲同定手法は、(a)ビート情報取得、(b)クロマベクトル取得、(c)類似度計算の 3 つからなる。今回の研究で筆者は(a)ビート情報取得の部分に焦点を当て、精度向上を図った。そのため(b)クロマベクトル取得、(c)類似度計算の 2 つの処理に関してはまだ手がつけられていない。特に実環境ライブバージョンに関して述べると、実環境ライブバージョンに含まれる実環境特有の成分はビートトラッキングだけでなくクロマベクトルに対しても悪影響を与えていると考えられる。以上より、今後は(b)と(c)の 2 つの処理に対して検討し、更なる同定精度の向上を目指したい。

#### 7. 参考文献

- [1] Shazam, <http://www.shazam.com>
- [2] Sound Hound, <http://www.soundhound.com/>
- [3] 石倉 和将ほか, “実環境におけるライブ演奏楽曲の楽曲同定”, 電子情報通信学会総合大会 2012.
- [4] Music Information Retrieval Evaluation eXchange, [http://www.musix-ir.org/mirex/wiki/MIREX\\_HOME](http://www.musix-ir.org/mirex/wiki/MIREX_HOME)
- [5] Daniel P.W. Ellis, “Identifying ‘Cover Songs’ with Beat-Synchronous Chroma Features”, MIREX 2006.
- [6] Daniel P.W. Ellis, “Beat Tracking with Dynamic Programming”, MIREX 2006.
- [7] Suman Ravuri, et al, “Automatic Cover Song Detection: Moving from High Scores to General Classification”, MIREX 2009.
- [8] Teppo E. Ahonen, et al, “Compressing Quantized Tonal Centroid Vectors for Cover Song Identification”, MIREX 2011.
- [9] Matthew Riley, et al, “A Text Retrieval Approach To Content-Based Audio Retrieval”, ISMIR 2008.
- [10] Daniel P.W. Ellis, “Beat Tracking by Dynamic Programming”, J. New Music Research, 2007.