

音楽音響信号に対する ビートトラッキングのための拍認識

鈴木 啓高 斉藤 博昭 中西 正和

慶應義塾大学大学院 理工学研究科 計算機科学専攻

1. はじめに

音楽情報処理の研究において音楽認識の分野では、対象を MIDI データに限定したものが中心であった。最近になって対象を一般の音楽音響信号にまで広げた研究がいくつか報告されているが、一般の音楽音響信号を対象とする場合、従来の MIDI データの場合とは異なる問題が発生するため、実用的なシステムの構築が困難である。本稿では音楽認識を実現する際に重要となる拍認識を実現する手法を提案する。

2. ビートトラッキング

西洋音楽は時間軸方向に、拍 (beat, ビート) < 小節 < 楽句 < 楽節 < 楽曲 という階層構造を形成する [4]。ビートトラッキングとは、入力となる音楽から階層構造を抽出する処理過程であると定義する [3]。これは、人間が音楽に合わせて手拍子を打ったり、歌ったり、演奏したりする際に不可欠な処理過程であると考えられる。

後藤らは一般の音楽音響信号を入力とし、そこに含まれる拍構造を認識するシステムを分散メモリ型の並列計算機である富士通 AP1000 上に実装し、その有効性を示した。

3. 大局的視野からの拍認識

これまでの研究では、入力の音楽音響信号を詳細に解析し得られる情報を元に拍を予測していくというのが中心であった。音楽的知識を持たない人でもある程度の拍認識を行なうことが可能であるという事実から、音程同定等の詳細な認識を行なうことなく拍の位置を抽出することが可能であると考えられる。本稿では音楽的知識のない人間が拍を認識する過程を

Extraction of Temporal Position of Beats in Musical Audio Signals

Hiroataka SUZUKI Hiroaki SAITO Masakazu NAKANISHI
Department of Computer Science, Faculty of Science and
Technology, Keio University 3-14-1 Hiyoshi, Kohoku-ku,
Yokohama, Kanagawa 223, Japan

Tel: 045-563-1141(3765)

(Email: {hiro, hxs, czl}@nak.ics.keio.ac.jp)

参考にし、入力の音楽音響信号を大局的視野から解析し、拍認識を行なう手法を提案する。以下にその手法について述べる。

4. 拍候補の抽出

4.1 仮定の設定

拍の位置に関して以下に示すいくつかの簡単な仮定を行なう。

仮定 1: 拍は等間隔に表れる可能性が高い

仮定 2: 拍の位置は発音位置に表れる可能性が高い

仮定 3: 拍の位置でコードが変化する可能性が高い

仮定 4: 低域の楽音が拍を構成することが多い

これらの仮定を用いて拍の位置を抽出する際に、さらに以下に示す仮定を適用する。

仮定 5: 発音の有無を認識することが出来ることが多い

仮定 6: 音の高い低いの違いを付けられることが多い

仮定 7: 特定の音だけを認識することが出来ることが多い

以上の仮定をもとに、拍認識を行なう。入力の音楽音響信号に対して高速フーリエ変換を用いて周波数解析を行ない、得られるスペクトルに対して先の仮定を適用していく。仮定 4 より得られたスペクトルの特に中域および低域に注目し、それぞれについて拍の位置の候補を決定する。以下に仮定の適用の過程と、拍位置の候補の決定手法について述べる。

4.2 中域からの抽出

時刻 t における周波数 f のパワーを $p(t, f)$ とした時、式 (1) または式 (4) を満たす時刻 t を拍候補とする。

$$\begin{cases} p_{min}(t, f) > MIDTHRESHOLD \\ p_{max}(t, f) < 0 \end{cases} \quad (1)$$

$$p_{min}(t, f) = \min(p(t, f), p(t+1, f)) \quad (2)$$

$$p_{max}(t, f) = \max(p(t-1, f), p(t, f \pm 1)) \quad (3)$$

$$\begin{cases} pmins(t, f) > MIDTHRESHOLD \\ pmaxs(t, f) < 0 \end{cases} \quad (4)$$

$$pmins(t, f) = \min(p(t, f \pm 1), p(t, f + 2)) \quad (5)$$

$$pmaxs(t, f) = \max(p(t-1, f), p(t-1, f \pm 1)) \quad (6)$$

4.3 低域からの抽出

時刻 t における周波数 f のパワーを $p(t, f)$ とした時, 式 (7) を満たす時刻 t を拍候補とする.

$$\begin{cases} p(t, f) > BASSTHRESHOLD \\ pmaxb(t \pm 1, f) < p(t, f) \end{cases} \quad (7)$$

$$pmaxb(t, f) = \max(p(t, f), p(t, f \pm 1), p(t, f + 2)) \quad (8)$$

4.4 動的なテンプレート生成とマッチング

人間が音楽を聴く場合に, 1 曲の中で同じフレーズの繰り返しや, 同じリズムパターンの繰り返しを認識することができる. 以前に耳にした情報を保持し, 新しい入力情報とのマッチングを無意識に行なっていると考えられる. 「以前の情報と新しいものとのマッチングを行なう」という人間の行動をテンプレートマッチングを用いて実現する.

ある拍候補の位置を抽出した際に, その時刻における周波数スペクトルをテンプレートとして保持する. それ以降の拍候補を求める際にこのテンプレートを利用することで, 拍抽出の助けとする. 新しい拍候補が得られた際にはただちにテンプレートを更新し, それまでのテンプレートは破棄する (図 1).

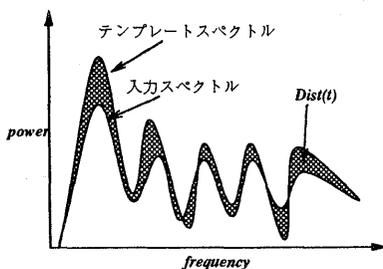


図 1: テンプレートマッチング

5. 拍候補の決定

抽出した拍候補が「確からしい」場合に, その時刻 t を拍の位置とする. 本稿では確信度を定義し, これを用いて確からしさを定める.

拍位置の候補の時刻を t とした場合, $T = t-1, t, t+1$ を満たす時刻 T において次の拍候補の抽出を行なう. $t-1, t, t+1$ の範囲内でのテンポのずれには追従することが可能である. 以上の処理を繰り返すことで入力音楽音響信号に追従していく.

6. 実験

本稿で提案した手法を用いて, 一般の音楽音響信号に対する拍認識を試みた. 実験にはドラムを含む一般的なポピュラーミュージックの楽器構成で, 楽曲を通してテンポがほぼ一定であるものを用いた. 実験の結果, 拍の位置をほぼ正しく認識出来ることが確認できた. ただし, 拍構造に関しては認識することができない. 4 分音符レベルでの認識, 8 分音符レベルでの認識といった区別を行なうことはできていない.

7. おわりに

本稿では音楽的知識のない人間が拍を認識する過程を参考にし, 入力音楽音響信号を大域的視野から解析し, 拍認識を実現する手法を提案し, その有効性を確認した. 今後は, さらに拍構造を解析していく手法を検討したい. また, 音楽情報処理の研究において特にその重要性が言われている評価方法の決定が今後の課題として挙げられる. より人間の主観による判定を反映する評価方法の確立が重要であると考え.

参考文献

- [1] Leigh M. Smith. Modelling Rhythm Perception by Continuous Time-Frequency Analysis, *In Proc. of ICMC*, pp.392-395, 1996.
- [2] Tim Kientzle. A Programmer's GUIDE TO SOUND, Addison-Wesley, 1997.
- [3] 後藤真孝, 村岡洋一. ビートトラッキングシステムの並列計算機への実装—AP1000 によるリアルタイム音楽情報処理—, 情報処理学会論文誌, Vol.37, No.7, pp.1460-1468, July, 1996.
- [4] 後藤真孝. 拍節認識 (ビートトラッキング), コンピュータと音楽の世界, pp.100-116, 共立出版, 1998.