動画の盛り上がり度に基づいたループシーケンサ

安坂 文汰[1,a)]　北原 鉄朗[1,b)]

概要：
これまで動画を用いた楽曲生成に関する研究は数多く行われてきたが，これらの多くは動画全体の色や動きから作曲を行うものであり，動画の盛り上がりを考慮した作曲は実現していない．本研究では，動画の盛り上がりに対応した楽曲生成を実現するために，動画の場面が持つ盛り上がりに基づいた楽曲の自動作成を提案する．動画に対して一定時間ごとに動きの特徴量を算出し，動画の盛り上がり度を推定する．推定した盛り上がり度と楽曲のセクションを考慮した隠れマルコフモデル（HMM）によって，ループシーケンサに挿入する音素材を決定する．以上の処理を実現したシステムの実装を行ったところ，楽曲構成を考慮しない場合に比べて，動画との対応やテクノミュージックとしての自然さに関して，概ね良い結果が得られた．

1. はじめに

近年，スマートフォンの普及により，Youtube やニコニコ動画などの動画共有サービスが人気を博している．これらの影響によって人々が動画を編集する機会が増加し，動画編集を支援するアプリケーションも普及した．より良い動画を作成するためには，バックグラウンドミュージック（BGM）の付与は必要不可欠だろう．しかしこのようなアプリケーションを用いても，使用する楽曲は自身で探す必要がある．使える楽曲には，著作権などの問題によって限りがある．それに加えて，素人が作曲を行うのは難しいため，楽曲を用意するのは困難である．仮に見つかったとしても，動画の長さに合うように調整する，といった手間や技術が必要となる．

こうした問題を解決しうる技術として，自動作曲技術がある．自動作曲についてはこれまで，歌詞の韻律に基づいて旋律を生成する Orpheus[1] やユーザが描いた盛り上がり曲線にあわせて楽曲を自動生成するループシーケンサ [2] がある．これらは歌詞や自身が描画した曲線に基づいて作曲を行っている．そのため，作成した楽曲をそのまま動画の BGM に付与するのは難しい．

一方，動画からの楽曲作成としては，動画の色や動きといった特徴から印象を推定し，楽曲を自動生成するものや [3] や動画の色彩と調との対応関係からの自動作曲 [4] があり，これらは動画全体を通して色や動きの特徴を印象に変換し楽曲を生成するものや，動画の色情報を調やコードと対応づけるものである．しかし，色などの動画特徴量と調やコードなどの音楽特徴量に対して適切な対応付けを定義するのは必ずしも簡単ではない．その他にスポーツビデオから場面分割を検出し，音楽を付与するシステム [5] がある．これは入力動画に合った BGM をデータベースの楽曲を用いて自動的に付加することができるが，楽曲の生成や動画の盛り上がりと楽曲の盛り上がりの対応が考慮されていない．映像の盛り上がりに関しては盛り上がり箇所に音楽のサビを同期させる研究 [6] がある．これは動画の盛り上がる場所の用意した楽曲のサビを付加するもので，一から楽曲を生成することは考慮されていない．

動画には，ストーリーやメッセージを伝えることを重視する場合の他に，躍動感，スピード感，スケール感，疾走感のような様々な印象を伝えることを重視する場合がある．実際，Adobe Stock[*1] のような動画素材サイトにおいて，これらのキーワードで素材を検索すると様々な素材が出力される．このような動画では，被写体の動き，カメラワーク，カット割りなどを工夫することで躍動感などを表現している．一方，躍動感やスピード感を重視する音楽ジャンルにテクノミュージック（以下，テクノ）がある．テクノはクラブミュージックの一種であり，比較的単純なドラムパターンを何度も繰り返すことでスピード感を出し，次々に音素材を加えていくことで楽曲としての盛り上がりやメリハリを表現する．

本稿では，躍動感やスピード感を重視して作られた動画

[1] 日本大学文理学部
[a)] yasusaka@kthrlab.jp
[b)] kitahara@kthrlab.jp

[*1] https://stock.adobe.com/jp/

にテクノの BGM を付与するシステムを提案する．テクノを作曲するツールの一つにループシーケンサがある．ループシーケンサは，通常のシーケンサのように音符を入力するのではなく，数秒の音素材を組み合わせて作曲するツールである．代表的なループシーケンサである「ACID Pro」[9] は数千を超える音素材を自由に組み合わせて曲を作ることができる．しかし，多数の音素材から適切なものを選ぶのは必ずしも簡単ではない．そこで，飯島らは，盛り上がり度の時間軌跡をマウスなどで入力させ，それに基づいて自動で音素材を選ぶことで，手軽に作曲できるようにしたシステムを提案した [2]．本稿で提案するシステムは，このシステムの拡張にあたり，動画中の躍動から盛り上がり度を求めて，その盛り上がり度から BGM を自動的に生成する．

2. 提案手法

本稿で提案するシステムは，上で述べたように、飯島らが開発したループシーケンサ [2] の拡張である．飯島らのループシーケンサは，盛り上がり度の時間軌跡をマウスで入力させ，その盛り上がり度に基づいて挿入する音素材の個数や種類を決めていた．本システムは，動画に含まれる躍動（動画中のオブジェクトや動画全体の動きの量）から盛り上がり度の時間軌跡を求め，そこからは飯島らのループシーケンサと同じ手法で楽曲を生成する．この手法には，動画は盛り上がるほど躍動が大きくなり，また，BGM においても盛り上がるほど音の数が増えたり，音が派手になる，という前提がある．

しかし，動画には動画の，テクノにはテクノの盛り上がり方があり，単に動画中の躍動に忠実に音素材を挿入しても，テクノとして適切なものになるとは限らない．そこで，テクノに関わりの深い Electronic Dance Music (EDM) の典型的な楽曲構成を参考に，楽曲構成に制限を加えることで，動画にある程度合わせつつもテクノとしての適切さを失わないようにする．

以下，動画から盛り上がり度を求め，そこから BGM を生成する手法を述べる．現状の実装では，用いる音素材はすでに 1 小節単位で提供されており，すべて同じテンポであることを前提としている．つまり，1 小節あたりの長さは一定とする．楽曲の長さは 4 小節の倍数であることが多いことを考慮し，動画の長さを超えない範囲で，できるだけ楽曲が長く小節数が 4 の倍数になるように小節数を決定する．以下，小節数を N とする．

2.1 動画からの盛り上がり度の抽出

まず，BGM を付与したい動画を読み込む．読み込んだ動画のフレーム間に対してモーションテンプレート解析を行う．「モーションテンプレート」は MIT Media Lab が開発した動画の動き抽出の効率的な方法である [7] [8]． 本手法では，各フレームの画像を複数の正方形のきまった幅にブロックに分割し，ブロックごとに，隣り合うフレーム間の動きの有無を検出する．動きの検出のあったブロック数を合計し，そのフレームにおける盛り上がり度とする．

以降の処理は小節単位で行うため，各小節に属するフレームの盛り上がり度の平均値を，その小節の盛り上がり度とする．以降，各小節の盛り上がり度を $x_0, \cdots, x_{N-1}$ とする．この方法で求められる盛り上がり度は，動画によってその値の幅が大きく異なるため，次式で正規化して 5 段階に離散化する:

$$x'_n = \mathrm{round}\left(\frac{4x_n}{\max(x_0, \cdots, x_{N-1})}\right) \tag{1}$$

ここで，x'_n が離散化後の盛り上がり度，max は最大値を求める関数，round は四捨五入を行う関数とする．

2.2 楽曲の構成の決定

上で述べたように，動画とテクノには本来異なるスタイルの盛り上がり方があり，単に動画の盛り上がりの軌跡に合わせて楽曲を生成するだけでは，テクノとして適切なものにはならない可能性がある．

Web ページ [10] によると，EDM は 4 個のセクションに分割することができる．楽曲の最初である「Intro」はペースを整え，ディスクジョッキー（DJ）が前の曲と繋ぐために使われるのが一般的である．「Intro」が終わると，「Breakdown」に移り，聴いている人へ期待感を与える．「Breakdown」は「Buildup」へつながり，徐々に盛り上がった後，メインである「Drop」へと遷移する．Drop は，楽曲の中で最大の盛り上がりとなる．一方，Web ページ [11] では，EDM は上記のセクションにおける「Drop」の後ろに，曲をつなぐための「Outro」を加えた 5 個のセクションから構成されると言われている．これらを考慮し，本稿では「Intro」,「Breakdown」,「Buildup」,「Drop」,「Outro」のセクションを使用する．

セクションの割当は 4 小節毎に行う．それぞれのセクションは想定されている盛り上がりが違うと考えられる．そのため，それぞれのセクションに対して標準的な盛り上がりを設定し，(1) 式で求めた盛り上がりとの差が最小になるように，セクションを割り当てる．なお「Buildup」は「Breakdown」の後ろから 2 小節に適用する．

前節の手法で求めた小節ごとの盛り上がり度を $x'_0, x'_1, \cdots, x'_{N-1}$ とする．いま，楽曲構成における各セクションの遷移は 4 小節単位で行われると仮定しているので，4 小節ごとの盛り上がり度を $x''_n = (x'_{4n} + x'_{4n+1} + x'_{4n+2} + x'_{4n+3})/4$ $(n = 0, \cdots, N/4 - 1)$ と定義する．また，各セクションの標準的な盛り上がり度を $y_{\mathrm{intro}}, y_{\mathrm{breakdown}}, y_{\mathrm{drop}}, y_{\mathrm{outro}}$ とする．2 つの時系列 $\{x''_0, \cdots, x''_{N/4-1}\}$ と $\{y_{\mathrm{intro}}, y_{\mathrm{breakdown}}, y_{\mathrm{drop}}, y_{\mathrm{outro}}\}$ に対して dynamic time warping (DTW) を適用することで，

セクションと小節の対応を決定する．なお，現在の実装では，$y_{\text{intro}} = 0.0$, $y_{\text{breakdown}} = 1.0$, $y_{\text{drop}} = 3.3$, $y_{\text{outro}} = 0.0$ としている．局所距離関数としては，単に差の絶対値を用いる．

2.3 音素材の挿入

飯島らのループシーケンサ [2] と同様の手法で音素材を挿入する．音素材として用いる「Sound PooL vol.1」[12] の仕様に基づき，音素材を「Sequence」,「Synth」,「Bass」,「Drums」の 4 パートに分け，各小節には各パート最大 1 つの音素材を挿入できるものとする．各音素材には 0~4 の 5 段階で盛り上がり度が付与されているとする．挿入する音素材の決定は次の 2 つのステップで行われる．

(1) 挿入するかどうかの決定.
(2) 挿入するなら何を挿入するかの決定.

2.3.1 挿入するかどうかの決定

各パートに音素材を挿入するか，しないかの 2 つの選択肢があるため，1 小節における音素材のパターンは 2^4 となり，これを状態とみなす．以降，各小節の状態を $s_0, \cdots, s_{N-1}$ とする．(1) 式で求めた盛り上がり度を出力信号とみなせば，隠れマルコフモデルにより解決することができる．初期確率（**表 1**）と状態遷移確率はセクション毎に設定し，状態遷移確率は初期確率をそれぞれの状態に設定した．出力確率（**表 2**）は，音素材が挿入されるパートが多い状態ほど，高い盛り上がり度に出力される様に設定した．例えば「Intro」では高い盛り上がりを観測しても，挿入する音素材の数は少なくなるように，「Drop」では観測された盛り上がり度が低くても挿入する音素材の数は多くなるように遷移確率と初期確率を設定している．

2.3.2 挿入するなら何を挿入するかの決定

挿入する音素材は楽曲に一貫性をもたせるため，4 小節毎にそれぞれのパートと盛り上がり度に対してランダムで選択する．$x'_0, x'_1, \cdots, x'_{N-1}$ と $s_0, s_1, \cdots, s_{N-1}$ をもとに対応した音素材を挿入する．つまり，n 小節目の音素材は，次のように決定される．n が 4 で割り切れる場合，上で求めた状態 s_n において音素材が挿入されることになったパートは，盛り上がり度が x'_n の音素材からランダムに選択される．n が 4 で割り切れない場合，直前の小節の音素材がそのまま使われる．

3. 実装および実行例

3 章で述べた手法の有用性を評価するため，本システムと比較用システムを使用した比較実験を行った．この比較実験により，動画の盛り上がりに対応した楽曲が動画の盛り上がりを考慮し，テクノとして自然な曲になっているか検証する．

表 1 セクション毎の初期確率. 表内の IT は Intro，BD は Breakdown，BU は Buildup，DR は Drop，OT は Outro を表す．状態は右から Drums，Bass，Synth，Sequence を表し，○の時挿入され，×の時挿入しないものとする．

状態	IT	BD	BU	DR	OT
××××	0.00	0.00	0.00	0.00	0.00
×××○	0.10	0.06	0.05	0.03	0.05
××○×	0.06	0.06	0.05	0.03	0.05
××○○	0.10	0.15	0.10	0.03	0.05
×○××	0.06	0.06	0.05	0.03	0.10
×○×○	0.06	0.06	0.05	0.03	0.05
×○○×	0.06	0.06	0.05	0.03	0.05
×○○○	0.06	0.10	0.20	0.03	0.05
○×××	0.06	0.06	0.05	0.03	0.15
○××○	0.06	0.06	0.05	0.03	0.05
○×○×	0.06	0.06	0.05	0.03	0.05
○×○○	0.06	0.06	0.05	0.10	0.05
○○××	0.06	0.06	0.05	0.03	0.15
○○×○	0.06	0.06	0.05	0.03	0.05
○○○×	0.06	0.06	0.05	0.20	0.05
○○○○	0.06	0.06	0.05	0.30	0.05

表 2 全セクションへ適用する出力確率. 表の横軸の数値は盛り上がり度を表す．状態は右から Drums，Bass，Synth，Sequence を表し，○の時挿入され，×の時挿入しないものとする．

状態	0	1	2	3	4
××××	0.00	0.00	0.00	0.00	0.00
×××○	0.40	0.20	0.20	0.10	0.10
××○×	0.39	0.21	0.20	0.10	0.10
××○○	0.25	0.30	0.25	0.10	0.10
×○××	0.39	0.21	0.20	0.10	0.10
×○×○	0.25	0.30	0.25	0.10	0.10
×○○×	0.25	0.30	0.25	0.10	0.10
×○○○	0.15	0.20	0.30	0.20	0.15
○×××	0.35	0.25	0.20	0.10	0.10
○××○	0.20	0.20	0.30	0.10	0.10
○×○×	0.20	0.20	0.35	0.15	0.10
○×○○	0.10	0.10	0.20	0.25	0.35
○○××	0.20	0.20	0.35	0.15	0.10
○○×○	0.10	0.10	0.20	0.25	0.35
○○○×	0.10	0.10	0.20	0.25	0.35
○○○○	0.05	0.05	0.25	0.25	0.40

3.1 評価方法

本稿で提案するシステムは python3 を用いて実装した．動画のモーションテンプレート解析には OpenCV を用い，音素材は前に述べたように「Sound PooL vol.1」[12] の素材を使用した．

実験用動画に対して，本システムと比較用システムそれぞれが楽曲を付与した動画を生成する．比較用システムは既存研究 [2] と同じくセクションを考慮しないものと，セクションは考慮するが，4 小節間で音素材を統一しないもの 2 つを使用し，それぞれを手法 1，手法 2 とする．セク

ションを考慮し，音素材を 4 小節間で固定する提案手法を手法 3 とする．使用する動画は Free Video clips[13] 上にある動画を実験用に 32 小節分の長さに調整したものである．

動画 1 つに対して，3 つの手法を用いて楽曲生成行い，Web で公開し，回答者には評価をしてもらう．回答者には作曲経験，テクノを聴くかどうか，ループシーケンサの使用経験を事前に回答してもらう．作曲経験は，

1. 全くない
2. ほとんどない
3. 作曲を試したことはある
4. 趣味としてたまに作曲する
5. 日常的に作曲し，作品を公開，演奏している

テクノを聴くかどうかは，

1. 全く聴かない
2. ほとんど聴かない
3. たまに聴いている
4. 毎日ではないが，普段から聴いている
5. 毎日のように聴いている

ループシーケンサの使用経験は，

1. 全くない
2. ほとんどない
3. ループシーケンサを使用して作曲を試したことはある
4. 趣味としてループシーケンサを使用して作曲している
5. 日常的に，ループシーケンサを使用して作品を作り公開している

のそろぞれ 5 段階で回答してもらう．その後，3 つ手法に対して次の 2 つ質問を行う．

Q1 動画の盛り上がりに対応しているか
Q2 テクノとして自然な楽曲か

それぞれの回答に対してそう思うかどうかを 7 段階で評価してもらう．加えて理由を記述してもらう．

3.2 生成例

ある動画に対して手法 1，手法 2，手法 3 を用いて BGM を生成した結果をそれぞれ**図 1**[*2]，**図 2**[*3]，**図 3**[*4]に示す．システム画面の横軸は時間軸を表し，1 ブロックが 1 小節である．縦軸は上部と下部に分かれており，上部は盛り上がり度，下部は挿入された音素材の状態を表している．

読み込んだ動画は人々が様々な場所を歩く様子を撮影したものであり，背景の動きが大きく盛り上がり度の推定に影響した．雑踏を歩く場面で高い盛り上がり度が見られ，逆に人の少ない場所や建物のみが映っている場面では低い盛り上がり度が見られた．

手法 1 では，推定された盛り上がり度が音素材の挿入される数がそのまま反映されるため，図 1 の 13 小節目で，動画が突然動きの少ない場面に変化するため音素材が，大きく減少した．第一著者が聴いたところ，音素材の遷移に一貫性がなく，テクノとして不自然な生成となった．

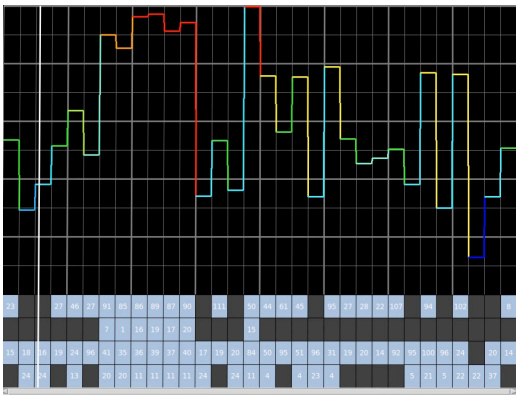

図 1 手法 1 を用いた盛り上がり度推定と音素材の挿入

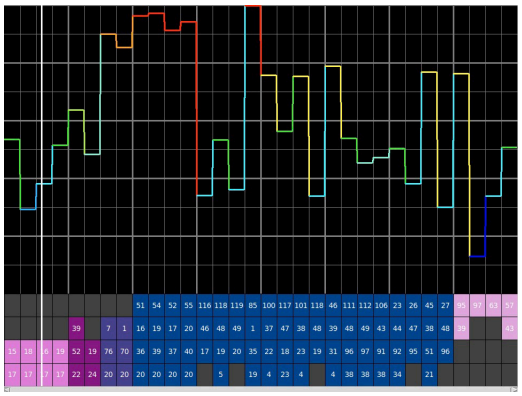

図 2 手法 2 を用いた盛り上がり度推定と音素材の挿入

手法 2 では，セクション分割を行い，それぞれに HMM を適用しているため，盛り上がり度の曲線とセクションの両方を考慮した音素材の挿入が行われた．手法 1 で問題だった 13 小節目ではセクションが「Drop」内であるため，HMM により音素材の数は減少するものの，大幅な減少ではなかった．第一著者が聴いたところ，音素材の状態数に大きな変化がないが，挿入される音素材に変化が多かった．これは，1 小節ごとにランダムに音素材が変わるからである．そのため，音素材の繋がりが不自然であった．

手法 3 では，手法 2 に加えて 4 小節間で音素材を揃えているため，推定された盛り上がり度の反映は少なくなったものの，13 小節目のような明らかに盛り上がり度の変化のあるところでは音素材の状態数が変化していた．第一著者が聴いたところ，音素材の挿入される数と挿入される音素材の変化ともに一貫性があり，3 つの手法の中で一番自然なテクノであると考える．

今回提案した手法では，4 小節ごとに 1 小節目の盛り上がり度に基づいて音素材の挿入を行ったが，4 小節ごとの平均値や最大値，最小値などを用いることもできる．何を用いるのが最も自然な生成が行えるかは，今後確かめていく必要がある．

[*2] 動画 1：https://youtu.be/bQWnn3UKiIc
[*3] 動画 2：https://youtu.be/mADGwJl_IAk
[*4] 動画 3：https://youtu.be/iytF0hWKeYE

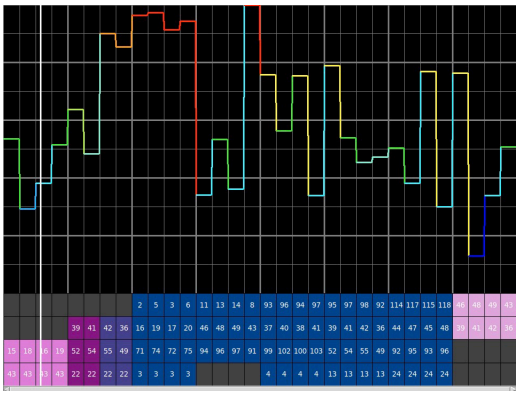

図 3　手法 3 を用いた盛り上がり度推定と音素材の挿入

3.3　評価結果と考察

上に述べた例を含んだ 3 つの動画を web 上に公開し，評価をしてもらった．風景が流れて途中で花火が上がる動画を動画 1，上に述べた動画を動画 2，色々な場所でスケートボードに乗っている動画を動画 3 とする．それぞれの動画における事前質問の回答を**表 3**，**表 4**，**表 5** に示す．3 つの手法で生成した楽曲に対する質問の回答を**表 6**，**表 7**，**表 8** に示し，それぞれの平均と標準偏差を求めた．動画 1 には 6 人，動画 2 には 6 人，動画 3 には 3 人の回答者が得られた．

動画 1 の結果 (表 6) において，手法 1 と手法 2 の結果を比較したところ，Q1，Q2 ともに手法 2 の方が評価の平均値が高く，標準偏差が小さい．この結果から生成される楽曲の構成に制約を設けることでテクノとして自然なものになっただけでなく動画との対応も改善されたと言える．また，手法 3 の評価の平均値が最も高く，標準偏差が低いため，手法 3 が 3 つの中で一番盛り上がりを考慮出来ていると言える．Q2 でも手法 3 の評価の平均値が最も高く，標準偏差は手法 2 より低いものの大きな差はなかったことから，手法 3 が 3 つの中で最も自然なテクノを生成できたとわかる．テクノをたまに聴く回答者からは「上の 2 つの動画よりも聴きやすく感じた」といったコメントもみられた．そのため，音素材を 4 小節で揃えることが自然なテクノを生成に繋がったと解釈できる．

動画 2 の結果（表 7）において，手法 1 と手法 2 の結果を比較したところ，Q1 では手法 1 が評価の平均がわずかに高く，手法 2 の標準偏差が低いため盛り上がりの推定においては手法 1 と手法 2 では余り差が見られなかった．Q2 では手法 2 の評価の平均が高いため，セクションを考慮することが自然なテクノを生成できると言える．また動画 1 と同様で手法 3 の Q1 の評価の平均値が最も高く，手法 3 が 3 つの中で一番盛り上がりを考慮出来ていると考えられる．Q2 では手法 3 の評価の標準偏差は一番高いが，平均値が大差で高いので，手法 3 が一番有用だと考えられる．実際に，ループシーケンサを使用したことがある回答者からは「統一感は出てきた」，その他の回答者からは「2 や 1

表 3　動画 1 の事前質問の回答

回答者	作曲経験	テクノを聴くかどうか	ループシーケンサの使用経験
1	1	1	1
2	1	2	1
3	1	1	1
4	1	1	1
5	1	1	1
6	1	2	2

表 4　動画 2 の事前質問の回答

回答者	作曲経験	テクノを聴くかどうか	ループシーケンサの使用経験
1	1	1	1
2	1	2	2
3	1	4	1
4	4	3	3
5	3	2	1
6	3	3	3

表 5　動画 3 の事前質問の回答

回答者	作曲経験	テクノを聴くかどうか	ループシーケンサの使用経験
1	1	2	1
2	1	3	1
3	1	1	1

より、より曲らしい曲だと感じた。急に音がキーキー鳴る印象が 2 より少なくて聴きやすい。急に来るのではなくだんだん盛り上げる感じ」といったコメントがあり，手法 1，手法 2 に比べて自然なテクノを生成したことがわかる．

動画 3 の結果（表 8）における，Q1 は動画 1，動画 2 と同様な結果となり，手法 3 に次いで，手法 2，手法 1 の順番に評価の平均が高く，標準偏差が低かった．Q2 では評価の平均値が 3 つの手法全て同じで，手法 1 と手法 3 は全員が同じ回答をした．そのため，この動画ではあまり自然なテクノが生成されたとは考えづらい．

しかしながら，今回は回答者が少なかったため，断定的な結論を下すことはできない．なかには，「盛り上がり点が自分の考えと違う」といったコメントもみられ，動画の特徴量として動き以外の要素を入れる必要がある他，盛り上がりという概念に対する個人間の考えの違いも検討する必要があると考えられる．

4.　おわりに

本稿では，動画の動きを特徴量とした盛り上がり度を推定し，盛り上がり度から楽曲の構成を自動で割り当てることで動画の盛り上がりに対応した楽曲を自動で生成する手

表 6 動画 1 に対する質問の回答

-	Q1			Q2		
回答者	手法 1	手法 2	手法 3	手法 1	手法 2	手法 3
1	7	7	7	5	6	7
2	2	2	3	4	4	3
3	7	7	7	7	4	7
4	1	5	6	3	3	7
5	3	4	4	2	3	4
6	5	7	6	5	7	6
平均	4.16	5.33	5.50	4.33	4.50	5.67
標準偏差	2.56	2.07	1.54	1.75	1.64	1.75

表 7 動画 2 に対する質問の回答

-	Q1			Q2		
回答者	手法 1	手法 2	手法 3	手法 1	手法 2	手法 3
1	6	3	5	3	4	6
2	2	5	5	3	5	4
3	6	4	4	3	5	7
4	6	6	6	1	2	3
5	5	5	6	3	5	6
6	2	3	3	3	5	6
平均	4.50	4.33	4.83	2.67	4.33	5.33
標準偏差	1.97	1.21	1.17	0.82	1.21	1.51

表 8 動画 3 に対する質問の回答

-	Q1			Q2		
回答者	手法 1	手法 2	手法 3	手法 1	手法 2	手法 3
1	6	6	6	4	4	4
2	6	7	6	5	6	5
3	2	5	6	5	4	5
平均	4.67	6.00	6.00	4.67	4.67	4.67
標準偏差	2.31	1.00	0.00	0.58	1.15	0.58

法を提案した．実際に動画を読み込み楽曲生成を行ったところ，セクション分割を行わない場合に比べて，行ったときは音素材の挿入数が大きく増減しない遷移をする楽曲が生成された．さらに 4 小節毎に音素材を揃えることでより自然なテクノが生成された．

今後，本システムの評価を回答者を増やして行うとともに，音素材の種類を増やし，様々な動画に対応した楽曲の自動生成を行いたい．今回はテクノに限定したが，その他の音楽ジャンルを持つ音素材にそれぞれ盛り上がり度をもたせることでテクノ以外の音楽を生成できるような拡張を予定している．その他，今回使用しなかった動画の色情報を分析することで，動画の盛り上がり度の推定方法の改善し，盛り上がり度の数を増やし，パート毎に挿入できる音素材の数を増やすことで，より多彩な楽曲を生成できるよう手法やシステムを拡張していきたい．

謝辞 本研究は，JSPS 科研費〈19K12288，16H01744，17H00749〉から支援を受けた．

参考文献

[1] 深山覚, 中妻啓, 米林裕一郎, 酒向慎司, 西本卓也, 小野順貴, 嵯峨山茂樹：Orpheus：歌詞の韻律に基づいた自動作曲システム, 情報処理学会研究報告音楽情報科学 (MUS), Vol.2008, No.78 (2008-MUS-076), pp.179–184 (2008)

[2] 飯島孔右, 鶴岡亜矢佳：手書き入力で盛り上がりをコントロールするループシーケンサ：スペクトログラムから盛り上がりの自動割り振り, 第 77 回全国大会講演論文集, Vol.2015, No.1, pp.373–374 (2015).

[3] 清水柚里奈, 菅野沙也, 伊藤貴之, 嵯峨山茂樹, 高塚正浩：動画解析・印象推定による動画 BGM の自動生成, 研究報告音楽情報科学 (MUS), Vol.2015, No.17, pp.1–6 (2015).

[4] 山本敏生，宝珍輝尚，野宮浩輝，動画をもとにした自動作曲，平成 21 年度情報処理学会関西支部大会論文集, vol.2009 (2009).

[5] J. Wang, E. Chng, C. Xu, H. Lu, Q. Tian: Generation of Personalized Music Sports Video Using Multimodal Cues, IEEE Transactions on Multimedia, Vol.9, No.3, pp.576–588 (2007).

[6] 佐藤晴紀，平井辰典，中野倫靖，後藤真孝，森島繁生，映像の盛り上がり箇所に音楽のサビを同期させる BGM 付加支援手法, 研究報告エンタテインメントコンピューティング (EC), Vol.2015, No.10, pp.1–6 (2015).

[7] Davis, James W and Bobick, Aaron F：The representation and recognition of action using temporal templates, IEEE conference on computer vision and pattern recognition, pp.928–934 (1997)

[8] Davis, James and Bradski, Gary：Real-time Motion Template Gradients using Intel CVLib, IEEE ICCV Workshop on Framerate Vision, pp.1–20 (1999)

[9] ACID Pro, 開発：MAGIX Software GmbH, 販売・サポート：ソースネクスト株式会社, 入手先 ⟨https://www.sourcenext.com/product/vegas/acidpro/⟩

[10] EDM Song Structure：Turn Your Loop Into A Song!-Cymatics.fm, 入手先 ⟨https://cymatics.fm/blogs/production/edm-song-structure⟩

[11] EDM の構成と作り方 | Madison Mars Milky Way(online), 入手先 ⟨https://salondemuze.com/blog/tip/edm-structure/⟩

[12] Sound PooL | 製品情報 | AHS(AH-Software)(online), 入手先 ⟨https://www.ah-soft.com/soundpool/⟩

[13] Free video clips(online), 入手先 ⟨https://mazwai.com/⟩