DanceReProducer:
既存のダンス動画の再利用により音楽に合った動画を作成できるシステム

DanceReProducer: A Music Video Creation System by Reusing Dance Video Content

室伏 空　　中野 倫靖　　後藤 真孝　　森島 繁生*

Summary.

本研究では，既存のダンス動画コンテンツの動画像を分割・伸縮して連結（切り貼り）することで，音楽に合ったダンス動画を自動生成し，ユーザの好みに合うようにインタラクティブに編集可能なシステム DanceReProducer を提案する．これによってユーザは，単に音楽を聴取するだけでなく，音楽に合った好みの映像を容易に作成でき，視覚的にも楽しむことができる．従来，音楽に合わせた映像付与に関する研究はあったが，それらは既存のダンス動画コンテンツを再利用して，音楽に合った動画を作成することはできなかった．本システムでは，Web 上で公開されている大量の二次創作動画を利用して，映像と音楽の対応関係をモデル化し，それに基づいて音楽に合った映像を自動的に作成する．さらに，自動的に作成された映像にユーザが「ダメ出し」するだけで，好みに合わない映像を容易に訂正・編集できる．本システムを実装して運用した結果，音楽に合わせたダンス動画の作成が効率的に行えて，また二次創作の経験の無いユーザであっても，音楽に合った映像を作成することができた．その際，自動生成された映像が好みのものでなくても,「ダメ出し」機能を使用するだけで容易に映像を訂正できることを確認した．

1　はじめに

近年，既存の動画像を断片的に利用（切り貼り）して，音楽に合わせた動画を作成して楽しむといった，新しい音楽鑑賞の文化が形成されつつある．既存の動画を一次創作とすると，それらを素材として別の様々なユーザが，それらの動画を音楽に合わせて切り貼りして作成したこのような動画は，二次創作に位置付けられる．これらの二次創作された動画は MAD 動画と呼ばれ，特に，リズム・印象・文脈的な意味等が，音楽と映像で高度に対応付けられた動画は視聴者から高い評価を受け，Web 上の動画共有サイト等で数多く再生されている．

このように，元々は違う音楽のために作られた映像を切り貼りすることで，音楽鑑賞に視覚的な楽しみを加えるだけでなく，好みの音楽に合った好みの映像を付与して新たな動画を楽しむことができる．しかし，このような二次創作動画を作成しようとした場合，音楽のリズムや印象に合った既存動画を探し出して，それを切り貼りして音楽のリズム・テンポに合うように伸縮するといった，高度で手間隙のかかる作業が必要とされる．さらに，音楽の文脈に沿って動画を構成するためには，音楽の構造や文脈を耳で聴いて判断するしかなく，それが動画制作の作業効率を下げていた．また，視聴を主な目的とした人にとっては，好みの楽曲を映像付きで楽しみたかったとしても，別の誰かがその曲の MAD 動画を作ってくれるのを待つしかなかった．

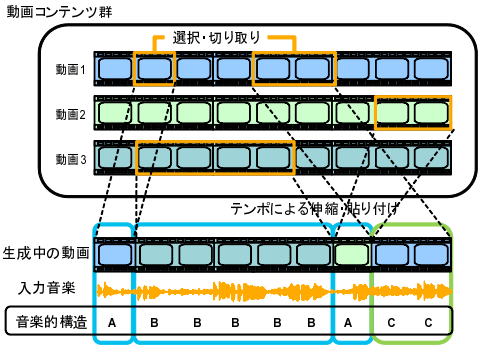

図 1. ダンス動画コンテンツを再利用して音楽に合った動画を自動生成するシステムのイメージ

従来，音楽に合わせた映像付与に関する先行事例がいくつか存在する．例えば近年の音楽再生ソフトウェアの多くには，音楽の周波数成分や強弱に反応した視覚効果を描画する機能がある．また様々な色や形（視覚効果）を描画する研究 [1] や，CG ダンサーなどのキャラクタモデルを音楽に同期させて動かす研究 [2, 3]，撮った映像を音楽に合わせて切り貼りすることで，ホームビデオを自動生成する研究等 [4, 5] があった．しかし，これらの方法では，既存の動画を再利用して音楽に合わせた動画を自動生成することを支援できなかった．

そこで本研究では，既存の動画コンテンツを利用

* Sora Murofushi and Shigeo Morishima, 早稲田大学, Tomoyasu Nakano and Masataka Goto, 産業技術総合研究所

して，音楽に合った映像を容易に作成できるシステム（図 1 参照）の構築を目指す．その第一段階として，音楽を入力として与えると，既存のダンス動画コンテンツの映像を切り貼りし，音楽に合ったダンス動画を容易に作成できるシステム「DanceReProducer」を提案する．ユーザは好みのダンス動画コンテンツをデータベースとして与え，それらを再利用してダンス動画コンテンツを作成できる．ここで，データベースには一次創作動画だけでなく，二次創作（MAD）動画も含め，それらから二次創作ないし三次創作動画を自動生成する．さらに，自動生成された結果が好みに合わない場合には，インタフェースを通じて「ダメ出し」を行うことでインタラクティブに編集できる，**ダメ出しインタラクション**を導入する．

2 DanceReProducer の設計

本章では，まず音楽に合った映像に対する考察を述べ，次にそれに基づいたシステムの設計方針と，インタフェースの機能について述べる．なお本論文では説明の便宜上，音楽付きのダンス動画コンテンツは単に**動画**と呼び，楽曲付きかそうでないかを区別するために，動画の映像部分を**映像**と呼ぶ．

2.1 音楽に合った映像についての考察

システムの設計にあたり，音楽に映像が合っていると感じられる要素を，以下に示す「局所的な対応関係」と「文脈的な対応関係」の二つの観点から考える．これは，動画の自動生成に関する従来研究 [4,5] や，二次創作コンテンツの制作者が Web 上などで公開している制作過程を参考にしたものである．

局所的な対応関係 音楽と映像の印象が合っていると感じる要素

- リズム：ダンス動作，カメラワークや画面の切り替え等による映像のリズムが，音楽のリズムやアクセントと同期している．
- 印　象：ダンス動作，映像全体の色彩や明るさ，ライティング等の映像効果の印象が音楽の印象と合っている．

文脈的な対応関係 音楽と映像の文脈的な流れが合っていると感じる要素

- 音楽的構造：音楽的構造（A メロ，B メロ，サビ等）に合わせて映像の印象が変化する．音楽の盛り上がりに合わせて映像が盛り上がる．
- 時間的連続性：音楽的まとまりの境界では映像シーンが切り替わり，それ以外の箇所では映像の印象が連続している．

以上は常に満たされるわけではなく，分類が相互に独立してはいないが，これらを考慮することで音楽に合った映像が作成できると考えられる．

2.2 映像の作成方法

2.1 節で述べた対応関係を満たす動画を作成するためには，音楽のリズムや印象に合った映像を既存の動画群から探し出し，音楽的な文脈に沿って映像を構成する必要がある．しかし，既存の動画群から音楽に合った映像を探し出すことは，仮に動画数が数十程度であっても，その一部の断片が切り貼り映像候補となるために膨大な候補数が存在し，人手で絞り込むのは容易でなかった．また，音楽の印象に合わせて映像を切り出して，伸縮してリズムを合わせる必要があった．さらに，候補映像を繋ぎ合わせて映像を構成するためには，音楽を何度も聴きながら，その構造や文脈を把握する必要があった．

そこで本システムでは，そのような作業を効率化するために，まず入力音楽に合った映像を自動生成する．しかし，自動選出された候補は必ずしもユーザにとって好みの映像とは限らない．そこで気に入らなかった場合には，ユーザが膨大な候補から選ぶことなく，インタフェース上でシステムに「ダメ出し」をするだけで容易に訂正可能にした．

本節ではシステムによる映像の自動生成の概要と，インタフェースの機能について述べる．

2.2.1 映像の自動生成

映像を自動生成するために，まずデータベース中の動画を音楽の小節単位で切り出して，局所的な対応関係へ対処する．これ以後，小節単位で分割された動画を**動画素片**，その集合を**動画素片集**と呼ぶ．音楽とそれに合わせたダンスは，曲のフレーズやダンス動作が楽曲のリズムに同期しているため，映像の切り貼りの最小単位として小節を用いる．

次に，入力音楽の各小節において，印象の合った動画素片の映像を選択する．それを入力音楽のテンポに合わせて伸縮させながら，繋ぎ合わせることで映像を自動生成する．この際，文脈的な対応関係を考慮するために，音楽的構造に基づいて，同一構造内では映像の印象が連続し（時間的連続性を保ち），構造の切り替わり箇所では印象を変えるように別の映像を選択する．

以上の処理を 2.1 節に対応させて次に示す．

リズムの同期 入力音楽の小節に合わせて動画素片の映像を時間方向に伸縮させることで，入力音楽とダンス映像のリズムの同期を実現する．

印象の近い映像選択 動画素片集の音楽と映像それぞれの印象に関する特徴量を自動抽出し，音楽と映像の対応関係をモデル化することで，入力音楽の印象にあった映像を自動選択する．

音楽的構造 音楽的構造の境界（自動推定したサビ区間や繰り返し区間の境界）で，映像の印象を変えることで，音楽的な文脈の流れに合わせた映像を作成する．

図 2. DanceReProducer のインタフェース画面

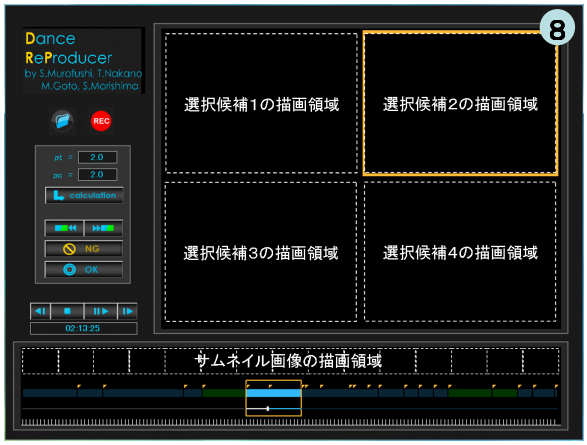

図 3. ダメ出しによって表示された他の候補

時間的連続性 映像を選択する際，時間的な前後関係を考慮し，不連続な接続にペナルティを与えることで，音楽的構造毎に映像の印象が連続するようにする．

以上のようにして，音楽と映像が局所的・文脈的に対応付けられた映像を自動生成する．

2.2.2 インタフェースによる視聴及び映像編集

図 2 に，本システムのインタフェース画面を示す．ユーザは本インタフェースを通じて，システムが自動生成した動画を視聴できるだけでなく，映像が気に入らなかった場合には，「ダメ出し」機能によって容易に映像を訂正・編集できる．

視聴のための基本機能として，現在時刻の映像の描画（図 2，①），音楽の読み込みや作成動画の保存（図 2，②），動画の再生・停止機能（図 2，③），再生時刻を表すスライダバーと音楽的構造の推定結果（図 2，④）がある．ここで構造の各区間はサビが緑，それ以外の区間が青で着色して表示される．また楽曲全体を通じて，時間的に等間隔な 15 箇所のサムネイル画像が描画される（図 2，⑤）．これらの機能を用い，自動生成された結果を確認できる．

また，好みに合わない映像に「ダメ出し」するために，インタフェースに以下の機能を与えた．

ダメ出し機能 NG ボタン（図 2，⑥）をクリックすることで，映像の描画領域を四分割して候補映像が描画される（図 3，⑧）．ユーザは四つの候補映像を同時再生して見比べ，よりイメージに合った映像を選択する．「ダメ出し」は音楽的構造の構造区間毎に行うが，これにより，音楽的な構造や文脈を考慮した映像を容易に作成できる．ここで，ユーザの好みの映像と出会う可能性を広げるため，映像候補には印象の異なる候補が提示されるようにした．

構造の頭出し機能 見通しよく編集するため，頭出しボタン（図 2，⑦）により音楽的構造の先頭へと移動ができる．また，構造区間（図 2，④）を直接クリックしても移動できる．

3 DanceReProducer の実現方法

本システムは，音楽と映像の対応関係をモデル化し，自動推定した音楽的構造と時間的連続性を考慮した映像の自動選出によって実現する．一般に音楽と映像との対応関係のモデル化は困難な課題だが，既存の二次創作動画群から学習することで実現する．二次創作動画には，同一の動画素材を再利用した様々な楽曲の動画や，逆に同一の楽曲に全く別の映像が対応付けられた動画が存在する．すなわち，人手で音楽と映像が対応付けられた大量の事例を入手して利用でき，音楽と映像に関する対応関係の多様性をモデル化できる可能性がある．このように音楽と映像との対応関係の多様な解釈をモデル化することは，学術的にも意義が深い．

システムは，既存の動画群から音楽と映像の特徴量を自動抽出して保存するデータベース構築フェーズと，局所的・文脈的な対応関係を考慮しながら映像選択を行う動画生成フェーズの二段階で構成される（図 4）．本章では，上記の観点を踏まえて，図 4 を参照しながら実装の詳細を述べる．

3.1 データベース構築フェーズ

データベース構築フェーズでは，Web 上から収集した動画コンテンツ群を次のように処理する．まず映像のフレームレート（fps）を 30 fps に，音楽のサンプリング周波数を 44.1kHz にリサンプリングする．これらから，1 フレーム（約 33 ms）毎に音楽と映像特徴量を抽出し（図 4，Ⓐ），フレーム特徴量と呼ぶ．次に，楽曲のテンポと小節線の位置を自動推定することで，フレーム特徴量を小節単位にまとめて小節特徴量を得る．以降，詳細を述べる．

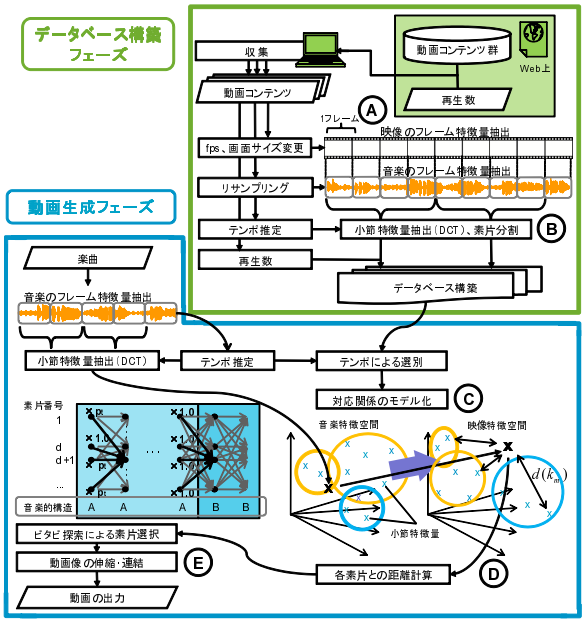

図 4. DanceReProducer の処理の流れ

3.1.1 楽曲のテンポ及び小節線の推定

楽曲のテンポ及び小節線の推定には様々な先行研究があり，将来的にはそうした成果を利用することも検討しているが，現段階では，予備実験において比較的良い結果が得られた，音響信号のパワーに基づく簡易的な方法で計算を行った．

まず，入力音響信号のパワーの自己相関関数のピーク時刻を求める．これはパワーの周期性を表すため，これを一拍の時間長としてテンポ推定する．ただし，倍テンポ誤りや半テンポ誤りを回避するため，テンポに 60〜120bpm（一拍が 0.5〜1.0 秒）の制限を設けて推定した．続いて，推定されたテンポから一拍毎の時刻にピークを持つパルス列を生成し，それと入力音響信号のパワーの相互相関関数を計算してピーク時刻を求める．これは楽曲中の一拍目の時刻を表すため，本論文では非常に単純な手法として，一拍目を小節線の開始位置とみなし，また 4/4 拍子を仮定して，機械的に小節線の位置を決定した．

3.1.2 音楽のフレーム特徴量抽出

音楽特徴量は，音楽と映像の対応関係に関する先行研究 [7,8] と，楽曲ジャンル分類に関する先行研究 [9] を参考に，アクセントおよび印象に関する特徴量を決定して抽出した．ここで，分析窓のシフト幅は映像特徴量と抽出間隔の対応を取るため，1470 点（= 44100 Hz ／ 30 fps）とした．

アクセントに関する特徴量としては，主に楽曲のパワーとその時間変化を表現するために，フィルタバンク毎のパワー（4 次元）と Spectral Flux（1 次元）を用いた．ここで，フィルタバンク出力はフィルタバンク数を 4 とした．印象に関する特徴量としては，楽曲の音色に関連した Zero-crossing rate（1 次元）と MFCC（Mel-Frequency Cepstral Coefficients）の直流成分と低次 12 項を用いた（13 次元）．

以上のように，音楽のアクセントと印象に関する計 19 次元のフレーム特徴量を抽出した．

3.1.3 映像のフレーム特徴量抽出

映像特徴量には，音楽と映像の対応付けに関する先行研究 [7,8] を参考に，アクセントおよび印象に関する特徴量を決定して抽出した．特徴抽出は，映像のフレームレートを 30 fps，画面サイズを 128×96 にリサンプリングして行った．

アクセントに関する特徴量としては，画面の動きやダンス動作とそれらの時間変化や画面の切り替わりを表現するために，オプティカルフローと輝度値の時間微分の平均値を用いた（各 1 次元）．オプティカルフローはブロックマッチング法を用い，ブロック数 64×48，シフト幅 1，最大シフト幅を 4 として計算した．印象に関する特徴量としては，映像の雰囲気を表現するために，全画素における色相，彩度，明度のそれぞれの値の平均と標準偏差を用いた（全 6 次元）．また映像全体の印象を表現するため，二次元の離散コサイン変換（DCT: Discrete Cosine Transform）の係数 12 次元（＝画面横軸方向の低次 4 項×画面縦軸方向の低次 3 項）を用いた．

以上のように，映像のアクセントと印象に関する計 20 次元のフレーム特徴量を抽出した．

3.1.4 小節特徴量の抽出

上述した音楽と映像のフレーム特徴量を小節特徴量としてまとめ，これを音楽と映像を対応付けるための特徴量として利用する．従来，楽曲のジャンル識別に関する研究等では，フレーム毎に抽出した特徴ベクトルを，各次元の平均と分散をとって楽曲の特徴量とすることが多かった [9] が，そのような方法では音楽と映像の時間方向の特徴が失われてしまう．

本論文では，時間的な変化を反映するために DCT を用いて小節特徴量にまとめる．ある小節のフレーム特徴量を，時間方向に 16 点へリサンプリングして各次元毎に DCT し，DCT 係数の低次 0 〜 3 項の計 4 次元を特徴量とした．すなわち，小節特徴量の次元数はフレーム特徴量の次元数の 4 倍となる．

3.2 動画生成フェーズ

動画生成フェーズでは動画素片集から映像を選択するために，次の処理を行う．

まず，データベース構築フェーズと同様に入力音楽から小節線の推定，小節特徴量の抽出，音楽的構造の推定を行う．続いて，音楽と映像の小節特徴量から，線形回帰によって音楽と映像の対応関係をモデル化し，局所的・文脈的な対応関係を考慮した映像選択を行って映像を自動生成する（図 4，Ⓒ – Ⓔ）．

ここで映像生成の際，不自然に速い（遅い）動作のダンス映像が生成されることを防ぐために，テンポが入力音楽と ±20% 以上異なる動画素片を選択候補から除外した．また，選別後の動画素片集について，それぞれの特徴量を主成分分析で次元削減して利用する（累積寄与率 95 %）．前節で述べた小節特徴量は，音楽特徴で 76 次元，映像特徴で 68 次元であったが，それぞれ 62 次元，68 次元へと削減された．ただし，入力音楽のテンポに応じて動画素片集を選別するため，次元削減数は若干変動する．

以降，線形回帰による音楽と映像の対応関係のモデル化と，局所的・文脈的な対応関係を考慮した映像選択について説明する．

3.2.1 複数の線形回帰に基づく音楽–映像の対応関係のモデル化

本論文では，音楽の小節特徴量を説明変数，映像の小節特徴量を目的変数とした線形回帰によって対応付けを行う．しかし，二次創作動画には，同一の映像素材が様々な楽曲に対応付けられた動画や，逆に全く別の映像素材が同一の楽曲に対応付けられた動画が存在し，単一の線形回帰のみ用いたのでは，そのような解釈の多様性を適切にモデル化できない．

そこで解決法として，複数の線形回帰モデルによって対応付けを行う．そのために，動画素片集の音楽と映像の小節特徴量を結合したベクトルをクラスタリングし，各クラスタ毎に線形回帰を学習した．ここで，クラスタリングには k-means 法を用いた．

3.2.2 局所的な対応関係と文脈的な対応関係を考慮した映像選択

局所的及び文脈的な対応関係を考慮して映像選択するために，入力音楽の小節毎にコストを求めて，それが入力音楽全体で最小となるように映像選択する．

局所的な対応関係は，小節毎に入力音楽から推定した映像特徴量と，動画素片集の全映像の映像特徴量との距離をコストとして計算する．入力音楽に対応する映像特徴量は，前項で得た回帰モデルを用いて，入力音楽の小節特徴量を映像の小節特徴空間へ写像して推定した．ここで回帰モデルには，入力音楽の小節特徴量に距離が最も近い重心を持つクラスタの線形回帰モデルを用いた．

また，文脈的な対応関係を考慮するために，生成される動画の時間的連続性や音楽的構造もコストとして考慮することで，動画像の生成をビタビ探索によるコスト最小化問題として解いた．ここで，音楽的構造とサビ区間は RefraiD [6] で求め，繰り返し区間の始端と終端を音楽的構造が切り替わる時刻として用いた．また，推定された繰り返し区間のうち，長さが四小節に満たないものは利用しなかった．

小節数が N の入力音楽について，小節番号を $n(n = 1, 2, ..., N)$，データベースの M 番目の動画の k 番目の動画素片を $k_m(k = 1, 2, ..., K_m, m \in M)$ で表した．動画素片毎のコストをユークリッド距離 $d(n, k_m)$ として，選択されるローカルコスト $c_l(n, k_m)$ と累積コスト $c_a(n, k_m)$ を次式で定義した．

$$c_l(n, k_m) = \begin{cases} d(n, k_m) & \text{if } ch(n) = ch(k_m) \\ p_c \times d(n, k_m) & \text{otherwise} \end{cases}, \quad (1)$$

$$c_a(n, k_m) = \min_{\tau,\mu} \begin{cases} c_l(n, k_m) & \text{if } \mu = m, \kappa = k - 1 \\ \quad +c_a(n-1, \kappa_\mu) & \vee st(n) \neq st(n-1) \\ p_t \times c_l(n, k_m) & \\ \quad +c_a(n-1, \kappa_\mu) & \text{otherwise} \end{cases} \quad (2)$$

ここで $ch(n)$ は小節 n がサビの場合に 1 を返し，$st(n)$ は音楽的構造の番号を返す関数である．また p_c が高いほど，データベース中の動画でサビに使われた映像が，入力音楽のサビで選択され易くなり，p_t が低いほど選出される映像が頻繁に切り替わる．

累積コストを最小化する映像系列は，最終小節 N において最も累積コストが小さい映像 $d_{\min}$ を次式で求めたのち，バックトレースによって得る．

$$d_{\min} = argmin_{k,m} c_a(N, k_m) \quad (3)$$

2.2.2 項で述べた「ダメ出し」機能では，選択された構造区間の最終小節において，累積コストが異なる四つの候補からバックトレースして映像を生成して表示する．生成される映像の幅を広げるために，最も累積コストが小さい候補，全候補数の 1/3 番目と 2/3 番目に累積コストが小さい候補，最も累積コストが大きい候補，の四つを選んで映像を生成した．これによって，多様な印象を持つ映像が生成できる．

3.3 再生数に応じた回帰モデルの学習

本研究では，Web 上に公開された二次創作動画を再利用しているが，これは次の問題を含むと考えられる．動画制作者は音楽と映像の対応関係を各々で解釈して動画を作成するため，音楽と映像の対応関係の信頼度，すなわち「映像が音楽と合っていると感じられる度合い」には，ばらつきがあると考えられる．したがって，全ての動画を均等に用いて対応付けを行うと，適切に学習できない可能性がある．

この問題を解決するために，動画共有サイトにおける再生数が，楽曲と映像の対応関係の信頼度を間接的に反映していると仮定し，対応関係の学習で利用する．具体的には，再生数に応じた重み付けのモデル学習を行った．再生数 V_c の動画は，モデル学習時に以下の式によって求まる重み w を用いる．

$$w = \alpha \times \lfloor \log_{10}(V_c) + 0.5 \rfloor + \beta. \quad (4)$$

ここで $\lfloor \cdot \rfloor$ は切捨を表し，0.5 を足して四捨五入する．

なお本研究では，$\alpha = 2$，$\beta = -7$ とした．すなわち 1 万回再生された動画は $w = 1$，10 万回再生された動画は $w = 3$ となるよう重み付けを行った．

4 システムの運用結果

本章では DanceReProducer 運用に利用した動画コンテンツと，運用の結果について述べる．

4.1 収集した動画コンテンツについて

本研究では，既存のダンス動画から音楽に合ったダンス動画を切り貼りして生成し，また，音楽からの動画生成に関する多様な対応関係をモデル化するために，システムが扱う動画コンテンツは以下に示す三つの条件を満たす必要がある．

条件 1 内容がダンスを中心に構成されていること
条件 2 動画を切り貼りして生成された動画であり，その素材が統制されていること
条件 3 上記二つの条件を満たすコンテンツが大量に存在し，かつ，入手が容易であること

このような条件を全て満たすコンテンツとして，バンダイナムコゲームスから販売されているアイドル育成シミュレーションゲーム「THE IDOLM@STER」とそのライブシミュレーションゲーム「アイドルマスター Live for You ! 」[10] を素材として二次創作された Web 上の動画を対象とした．ここで，再生数を対応関係の学習等に用いて動画生成を行うことを考慮し，再生数がカウントされている必要がある．そこで，動画共有サイト「ニコニコ動画」[11] から，再生数が 1 万回以上の動画を 100 件収集した．

4.2 システムの運用結果

本システムにより作成された動画は，リズムの同期や印象の近い映像が作成されていた．したがって，既存の動画群における音楽と映像の対応関係を適切にモデル化することができたといえる．

インタフェースを用いた実際の運用では，音楽と合っていない動画が自動生成される場合があっても，「ダメ出し」機能を使用するだけで映像が訂正できるため，好みの動画を簡単に作成できた．一方システムの改善点としては,「ダメ出し」による訂正の際に，映像の印象がほとんど類似した候補ばかりが提示されてしまって，適切に訂正できない場合もあった．

また，二次創作経験の無いユーザからの意見としては，映像候補数を増やすことで，イメージしやすくなるという意見を得た（現状では四候補）．二次創作経験のあるユーザからは，音楽の小節線や音楽的構造の推定結果を訂正できれば，より良い動画を作れそうだという意見を得た．

5 おわりに

本論文では，既存のダンス動画を再利用して音楽に合った動画を作成できるシステム「DanceReProducer」を提案した．これにより，音楽に合った動画が自動生成できるだけでなく,「ダメ出し」機能で音楽と印象の合わない映像を訂正して，音楽に合った映像を作成することが容易となった．技術的な成果としては，大量の二次創作動画における音楽と映像の多様な対応関係を，複数の線形回帰によってモデル化することができた．その際，動画の再生数を，回帰学習時の重みとして利用した．

DanceReProducer は，二次・三次創作を含む「N 次創作」[12] を支援できる．最近では，このような文化に適合する新しいコンテンツビジネスに関して，興味深い考察がなされている [12]．Web 上の素材を再利用できる支援技術を実現することで，そのような文化・産業におけるユーザ支援を考えていきたい．

システムの定量的な評価は今後の課題である．また，動画作成を支援するために，小節線の推定位置や音楽的構造の推定結果等を訂正できるインタフェースや,「ダメ出し」で提示する映像候補にバリエーションを加えるため，ダンスの動きを特徴量として抽出する手法などを検討する予定である．

参考文献

[1] 藤澤 隆史, 谷 光彬, 長田 典子, 片寄 晴弘. 和音性の定量的評価モデルに基づいた楽曲ムードの色彩表現インタフェース. 情報処理学会論文誌. pp. 1133–1138, 2009.

[2] M. Goto. An Audio-based Real-time Beat Tracking System for Music With or Without Drum-sounds. Journal of New Music Research, pp. 159–171, 2001.

[3] 白鳥 貴亮, 中澤 篤志, 池内 克史. 音楽特徴を考慮した舞踊動作の自動生成. 電子情報通信学会論文誌 D, pp. 2242–2252, 2007.

[4] J. Foote, M. Cooperand and A. Girgensohn. Creating music videos using automatic media analysis. Proceedings of the tenth ACM international conference on Multimedia, pp. 553–560, 2002.

[5] X.-S. Hua, L. Lu and H.-J. Zhang. Automatic music video generation based on temporal pattern analysis. Proceedings of the 12th annual ACM international conference on Multimedia, pp. 472–475, 2004.

[6] M. Goto. A Chorus-Section Detection Method for Musical Audio Signals and Its Application to a Music. IEEE Transactions on Audio, Speech, and Language Processing, pp. 1784–1794, 2006.

[7] O. Gillet and G. Richard. Comparing Audio and Video Segmentations for Music Videos Indexing. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, pp. V-21–V-24, 2006.

[8] 西山 正紘, 北原 鉄朗, 駒谷 和範, 尾形 哲也, 奥乃 博. マルチメディアコンテンツにおける音楽と映像の調和度計算モデル. 情報処理学会研究報告 2007-MUS-069, pp. 111–118, 2007.

[9] G. Tzanetakis and P. Cook. Musical Genre Classification of Audio Signals. IEEE Transactions on Speech and Audio Processing, pp. 293–302, 2002.

[10] バンダイナムコゲームス. THE IDOLM@STER OFFICIAL WEB. http://www.bandainamcogames.co.jp/cs/list/idolmaster/.

[11] ニワンゴ. ニコニコ動画 http://www.nicovideo.jp/.

[12] 濱野 智史. インターネット関連産業. デジタルコンテンツ白書 2009, pp.118–124, 2009.