Vol.2011-MUS-89 No.15
2011/2/12

音楽と映像の同期手法に基づく ダンス動画生成システム

平井辰典† 大矢隼士† 長谷川裕記† 森島繁生†

本稿では，主観評価実験によって評価された音楽と映像の同期手法に基づいて，入力された音楽から，人間が同期していると感じるダンス動画を生成するシステムを提案する．本システムの土台となる音楽と映像の同期手法は，音楽のエネルギーを示す特徴量である RMS に対し，映像のアクセント（明滅や動きの速さなど）を付加するというものである．これは，本研究において主観評価実験により人が音楽と映像が「合っている」と感じると確かめられた同期手法である．本システムでの動画生成は，まずデータベースの構築として既存のダンス動画シーケンスから人物領域のみを抽出し，Optical flow の計算を行う．それに対し入力音楽を分割した素片の RMS の挙動に最も近い挙動を示す Optical flow のダンスシーケンスをデータベース中から選択し，それらの映像シーケンスの切り貼りを行うことで，音楽に最も同期しているダンス動画の生成を行うというものである．

Dance Video Creation System by Synchronization between Music and Video Features

TATSUNORI HIRAI† HAYATO OHYA†
YUKI HASEGAWA† and SHIGEO MORISHIMA†

In this paper, we propose dance video creation system by the synchronization between music and video feature which evaluated by human's subjective judgment experiment. The video which created from the system matches with the criterion of synchronization which human tend to feel the music and the pictures are synchronized. The criterion of synchronization is that when RMS energy of music matches to the accent of video, people tends to feel the music and pictures are synchronized. In movie creation of this system, first thing to do is to make a database from existent dance movie's information about dance. We acquired them from optical-flow of the movies. The process to create dance movie is to cut the pieces of pictures that optical-flow shows best correlation to the RMS of the input music, and connect them together.

1. はじめに

近年，インターネットの通信速度の向上や，PC のメモリの大容量化に伴い、ニコニコ動画[1]に代表されるような動画共有サイトがインターネット上で流行している．なかでも，ニコニコ動画で「音楽」のカテゴリに分類されている動画の数は，61 万動画あり全動画数約 530 万動画のうちの約 11.5 %を占めており,最も需要の高い動画のジャンルとなっている．さらにニコニコ動画の会員数は年々増えており，今まで視聴するのみに留まっていたユーザが自分で新たなコンテンツを制作する機会も増加している．そのようなユーザが生成したコンテンツを指して，消費者生成メディア，CGM（Consumer Generated Media），UGC（User Generated Content）という言葉も生まれている[2].

本研究では，一般に Music Video や Promotion Video と呼ばれる音楽を主体とした映像作品（以降，音楽動画と呼ぶ）を，クリエータが使うような専門的な映像編集ツールなどを使わずに生成するシステムを提案する。それにより，今までコンテンツを享受するに留まっていたユーザが自分からコンテンツを制作してみようと思えるような,簡単に高品質なコンテンツの制作を実現できるシステムを目指す．このような CGM 支援により，ユーザのメディアへの新たな楽しみ方を生みだし，誰もが手軽にコンテンツ生成を体験でき，クリエータとなりうる環境の実現が期待される．

本稿では，本研究で実現を目指す音楽動画生成の初期段階として映像の対象をダンスのみに絞る．その理由として，映像の内容や雰囲気などの意味的理解を反映したシステムは複雑なものとなってしまい，生成結果のどの部分が人間にとって音楽と映像が「合っている」という印象を与える要因となるかが明確ではないためである．

2. 音楽と映像の同期手法

音楽動画を生成する上で重要となるのは，音楽と映像の同期基準である．逆に，音楽と映像の同期基準さえ明確になっていれば，その基準に従うだけで動画を生成できる．音楽動画の制作現場では，音楽がすでにある状態で音楽に合うような映像を編集していくということが多いので，音楽に最も合っている映像を付加する方法について検討していく．

音楽と映像の同期を決める際の基準となる要因として，音楽にはテンポやビート，リズムなどのアクセント，映像には光の明滅やオブジェクトの動きの変化などといったアクセントが挙げられる．人は，音楽のテンポに合わせて手拍子を打ったり，ステージの照明が明滅したりすることで音楽に対する調和を感じ,「気持ちいい」心理状態（情緒的反応）になり，それらのアクセントと音楽のテンポがずれると人は違和感を覚えるといった研究結果が報告されている[3-6]．また，さらに詳細な音楽と映像の同期にまで言及するために，音楽と映像の時間軸上での調和である時間的調和だけでな

ⓒ 2011 Information Processing Society of Japan

く音楽と映像のムードの一致による意味的調和の両方を考慮した調和度計算手法に関する研究報告もされている[7].

本稿では，システムの第一段階として，ダンスという動的な映像対象のみに限定したシステムの構築をしていくため，音楽と映像のムードの一致による意味的な調和の要因をすべて排除したうえで，時間軸上での調和の実現を目指した同期手法について検討した.

ここで，楽曲のテンポに映像のアクセントを一致させると人が同期を感じるという報告を元に，さらに同期を感じるような同期手法について考える．テンポとは，楽曲における 4 分音符の長さを左右するものであり，その曲が 1 分間に 4 分音符を何回分弾ける速さかを示す BPM（Beats Per Minute）で表される．つまり，楽曲のテンポと映像のアクセントが一致しているというのは，音楽 1 小節に対し，4 つの点において映像のアクセントが付加されている状態を示している．これでは，楽曲の小節中での 4 分音符よりも細かい単位での音の変化への対応が十分にできているとは言えない．例えば，16 分音符でのピアノ演奏に対して，4 分音符のリズムで照明が明滅する場合を考えると，テンポは一致しているため違和感は覚えないが，より詳細な視点で見たときに,照明も 16 分音符のリズムまたは明と滅をそれぞれ 1 アクセントと見て 8 分音符のリズムで明滅していた方がより同期を感じるのではないかと考えた．そのような詳細構造を考慮するにはテンポだけでは不十分であると言える.

2.1 詳細な音楽構造まで考慮した音楽と映像の同期手法

音楽や映像において，両者が大きく変化する箇所をマッチングさせると、両者が変化を示す箇所が一致していない動画に比べて音楽と映像の同期の度合いは大きく向上する[8]．そこで，音楽の変化に合わせて映像を変化させることで，テンポで同期させるよりも詳細な音楽構造まで考慮された同期を実現させることを考えた.具体的には，音楽の時間的な変化を表す特徴量として，音のエネルギーを表す RMS（Root Mean Square）に対して，映像のアクセントとして，オブジェクトの輝度値，オブジェクトの動きの速さをそれぞれ対応させた．この同期基準に基づき，音楽のエネルギーの強さに応じて画面上の単純なオブジェクトの輝度値が変わる映像，オブジェクトが水平方向に動く動きの速さが変わる映像をそれぞれ生成した．これにより，音量が大きくエネルギーが強いところでは，オブジェクトが強く光ったり速く動いたりし，音量が小さいところではオブジェクトが弱く光ったり遅く動いたりするような映像ができた（**図 1**）.

これを，音楽のより詳細な構造まで考慮されている音楽と映像の同期手法として，従来から提案されている音楽のテンポに合わせて映像にアクセントを付加する同期手法との比較を主観評価実験により行った.

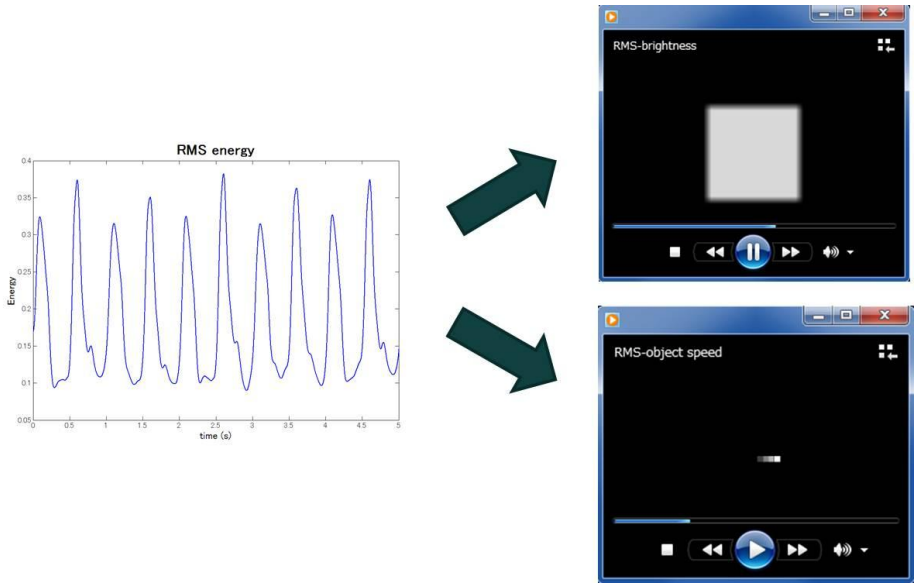

図 1　音楽の RMS に対応した映像の生成

2.2 主観評価実験

音楽の RMS に対して，映像のアクセント（映像の動きや明滅）を対応させることが，音楽のテンポと映像のアクセントを一致させる同期手法よりも，人がより「合っている」と感じる同期手法であるかという比較を主観評価実験により検証した.

主観評価実験は，音楽のテンポと映像のアクセントが一致するように生成した動画と,音楽の RMS に映像のアクセントが対応するように生成した動画とを AB 法で一対一比較することで行った．映像のアクセントとしては，映像の明滅の要素として，オブジェクト輝度値の大きさを音楽に対応させたもの，映像の動きの要素として，映像の速さ（前のフレームから次ぎのフレームに移る際のオブジェクトの移動距離）を音楽に対応させたもの，2 つの要素を組み合わせたものとして，オブジェクトの輝度値と速さ両方を音楽に対応させたものを使用した．また，オブジェクトは単純な正方形の物体とした.

音楽のテンポに合わせて映像のアクセントを付加する動画として，音楽の拍に合わせて明滅を繰り返す動画，音楽の拍に合わせて等速での動作と静止を繰り返す動画を生成した.

ⓒ 2011 Information Processing Society of Japan

主観評価実験は 20 代の男女 22 名に対して行った．本実験に使用した楽曲は，変拍子の楽曲を含む音楽的構造の異なる楽曲 6 曲とドラムによる単純なリズム音 2 パターンである．各楽曲に対して本同期手法と音楽のテンポに合わせる同期手法それぞれに基づいて映像のアクセントを付加させた動画を比較して，どちらがより「合っているか」を AB 法により，「A の動画の方が合っている」から「B の動画の方が合っている」までの 5 段階で評価させた（**表１**）（**表２**）．提案手法の方が合っている場合のスコアを 5 とし，どちらも同じくらいで合った場合を 3，従来の手法の方が合っている場合を 1 となるようにした．何を基準にして「合っている」と判断するかは個人差があるため，より「合っている」のはどちらの動画であるかのみを聞き，最後に内観調査として何を基準にして合っていると判断したかを回答させた．

表 1　主観評価実験

評価	スコア
A の動画の方が合っている	5
どちらかというと A の動画の方が合っている	4
どちらも同じくらい合っている	3
どちらかというと B の動画の方が合っている	2
B の動画の方が合っている	1

表 2　使用楽曲

曲名	BPM
Himitsu Girl’s Top Secret/Zazen Boys	139
Yureta Yureta Yureta/Zazen Boys	124
Dream Fighter/Perfume	135
Sugarless Girl/capsule	130
Space Party/自作曲	120
テルーの唄/手島葵	68
８ビートドラム音	120
変拍子ドラム音	120

2.3　実験結果

ほとんどの楽曲において，従来の音楽のテンポに合わせて映像のアクセントを付加する同期手法と同等またはそれ以上の評価が得られた（**図４**）（**表３**）．特に，リズムの変化があるような変拍子の楽曲では，本手法に沿って生成した動画の方が「より合っている」という結果が得られた．これは，変拍子の楽曲はリズムが一定の楽曲に比べて拍を取るのが困難であることによるものと考えられる．

楽曲「Dream Fighter」では，従来手法の方が「より合っている」という結果となったが，RMS の様子を観察してみると，この楽曲だけ推移の幅が極端に小さいことがわかった．これにより RMS の変化からのアクセントが十分に付加されてないことが原因であると考えられる．このような RMS の推移の幅が小さい楽曲の場合には，RMS の値を映像のアクセントに対応させる際にアクセントが十分に付加されるような処理を加える必要があると考えられる．

何を基準に合っていると判断したかという内観調査では，「映像が曲のテンポ/リズムに合っているか」というような回答をした人が 22 人中 10 人と最も多く，テンポとの一致の重要性も改めて確認された（**表４**）．

以上の結果から，音楽のエネルギーに映像のアクセントを一致させる同期手法は，人が「合っている」と感じる音楽と映像の同期手法であると言える．さらに，人間が音楽と映像の同期においてテンポを重要視していることも考えると，テンポを考慮した上で音楽の RMS と映像のアクセントとの同期をはかれば，音楽と映像が「より合っている」同期手法となると考えられる．

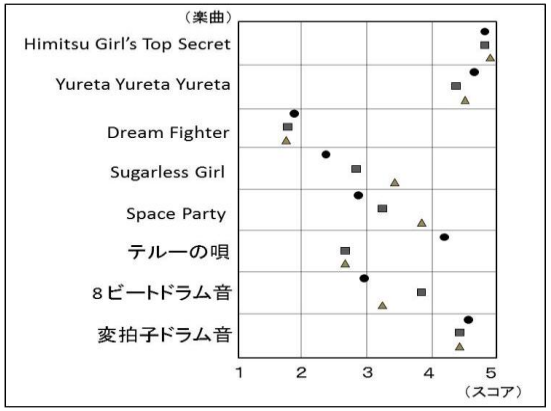

図 4　実験結果

ⓒ 2011 Information Processing Society of Japan

表 3　実験結果

曲名	映像アクセント	スコア
Himitsu Girl's Top Secret /Zazen Boys (変拍子楽曲)	明滅	4.82
	動き	4.82
	明滅＋動き	4.91
Yureta Yureta Yureta /Zazen Boys (変拍子楽曲)	明滅	4.64
	動き	4.36
	明滅＋動き	4.5
Dream Fighter /Perfume	明滅	1.86
	動き	1.77
	明滅＋動き	1.73
Sugarless Girl /capsule	明滅	2.36
	動き	2.82
	明滅＋動き	3.41
Space Party /自作曲	明滅	2.86
	動き	3.23
	明滅＋動き	3.82
テルーの唄/手島葵	明滅	4.18
	動き	2.64
	明滅＋動き	2.65
８ビートドラム音	明滅	2.95
	動き	3.82
	明滅＋動き	3.23
変拍子ドラム音	明滅	4.55
	動き	4.41
	明滅＋動き	4.41

表 4　内観調査

報告内容	報告者数
映像が曲のテンポ/リズムに合っている	10 名
映像が曲の雰囲気に合っている	6 名
映像がドラム音に合っている	5 名
映像がベース音に合っている	1 名

3. ダンス動画生成システム

本システムは，既存のダンス映像から人物の動きの情報のみを抽出したものをデータベース化するデータベース構築フェーズと，楽曲を入力としてそれに合ったダンス動画を生成する動画生成フェーズにより構成される．動き情報のみで表現できる映像の代表としてダンス映像のみに対象を絞り，既存のダンス映像を用いることで，動きの大きさの情報のみが一致する映像を選択するという単純な条件で動画の生成が実現できる．

3.1 システム設計

主観評価実験での音楽のエネルギーである RMS に対して映像のアクセントとして動きを対応させると人が「合っている」と感じるという結果に基づきシステムを構築する．データベース中のダンス動画の中から，入力楽曲の RMS の推移に最も近い推移を示すダンスの動き情報をもつ動画を探索し，該当動画の映像を入力楽曲に貼り付けることで新たなダンス動画の生成を行う．これを入力楽曲の任意の長さに対して繰り返し行うことで，既存のダンス動画の様々な区間を切り貼りしてできた新たなダンス動画が生成される．

3.2 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，既存のダンス動画群から，それぞれのダンスの人物の動きの特報量であるオプティカルフローを抽出する．本システムのオプティカルフローの抽出はブロックマッチング法を用いて行った．また，人物の動きの特徴のみを抽出するために，データベース化に用いる動画は背景が固定されていてカメラの切り替えなどのイベントが起こらないものを収集することが望ましい．また，複数の映像素片の切り貼りによって動画を生成するため，映像の切り替えの際の違和感をなくすためにデータベース中の動画はなるべく色相などの条件が近いものの集まりであることが望まれる．本システムでは，このような条件を満たす動画として，バンダイナムコゲームスから発売されている音楽ゲームであるアイドルマスター・ライブフォーユー[9]のプレイ画面で，ダンスの背景がブルーバックとなっているものをキャプチャ

ⓒ 2011 Information Processing Society of Japan

したものをデータベースとして用いた．背景をブルーバックとすることで，オプティカルフローでキャラクターのダンスの動きの情報のみを効率的に捕らえられる．

3.3 動画生成フェーズ

動画生成フェーズでは，入力楽曲の RMS の推移に応じて，データベースから最も「合っている」映像の素片を探索して，選ばれた動画同士を切り貼りすることで新たな動画を生成する．

入力楽曲の RMS は，動画のフレームレートに合わせ，1 秒間に 30 サンプルずつサンプリングする．さらに，入力楽曲のテンポ推定を元に楽曲の小節線を推定することで，映像を探索するための区間の長さを小節の切れ目間などに設定でき，より自然な間隔で映像の切り替えが行われる．

入力楽曲に最も「合っている」映像素片の探索は，RMS の推移とデータベース中の全オプティカルフローの推移との間で以下の式で表される相関係数 R を求め，最も大きな値を示すものを選択することで行われる．

$$R = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \quad . \tag{1}$$

ここで，x には楽曲の RMS を正規化したもの，y にはデータベース中のオプティカルフローを正規化したもの代入するものとする．相関係数は-1 から 1 の間の値をとり，0 に近いほど二つのデータ間に相関がなく，1 に近ければ両データの推移の間に線形の関係があり，変化の様子が近いということが言える．

以上のようにして生成されたダンス動画は，主観評価実験の結果を反映した，音楽の詳細な構造の変化に対応した映像の付加が行われた動画であると考えられる．また，本システムの性質上，データベースの動画数が多いほど，入力楽曲の RMS の推移に近いオプティカルフローの推移を示す動画の存在確率が高くなり，より主観評価実験の結果に即した動画が生成されやすくなる．

4. 生成結果と考察

本システムを用いて生成されたダンス動画は，RMS の推移に応じて映像中のキャラクターの動きの大きさが変わるという，主観評価実験の結果を反映した動画となった．そのため，音の変化に合わせてキャラクターが動いている映像が，音が鳴っていないところではキャラクターが静止している映像が選択されるなど，音楽と映像が同期している動画が生成できたと言える．

本システムの定量的な評価は今後の課題であり，本システムにより生成されたダンス動画が，テンポだけを考慮してランダムに切り貼りされて生成されたダンス動画よりも人間が見て「合っている」と感じるかどうかが重要である．それにより，テンポだけが一致していればいいのではなく，音の変化と動きの変化の様子も一致していた方がより「合っている」という主観評価実験結果に沿ったシステムとなっていることが確認できる．

5. おわりに

本論文では，主観評価実験を元に，人間が合っていると感じる音楽と映像の同期手法を提案し，その手法に基づいて既存のダンス動画シーケンス群の中から音楽に最も合った映像を選択して切り貼りすることで，新たなダンス動画生成を実現するシステムを提案した．

本システムの現状としては，入力楽曲により音楽と映像の同期の度合いに差があるが，音の変化に最も近いキャラクターの動きを示す映像を選択することで，主観評価実験の結果に近く，ある程度の同期を感じられるようなダンス動画が生成できた．今後データベースの拡充をしていくことにより，探している音にさらに適した映像素片が見つかりやすくなることが期待できる．

また，実際のダンスに関する知見なども考慮し，ダンスにおける音楽に対する自然な動きも考慮していきたい．

今後，データベースを拡充することにより様々な種類の動画に対してもシステムを対応させていく予定であり，それと同時に使用する映像の特徴量を増やしていくことで，より複雑な内容の映像にも対応できるようになる予定である．さらに，複雑な音楽と映像の対応関係について研究していき，音楽のジャンルや映像のジャンルにもとらわれない汎用的な動画生成システムを構築すること目指したい．また，映像の意味的理解をはかり，楽曲の歌詞などとの意味的調和などを目指すことで，人々が「おもしろい」と感じるコンテンツの生成を支援していきたい．

参考文献

1) ニワンゴ：ニコニコ動画, http://www.nicovideo.jp/.
2) 後藤真孝，剣持秀紀，伊藤博之，戀塚昭彦，濱野智史：「特別セッション：CGM の現在と未来:初音ミク，ニコニコ動画，ピアプロの切り拓いた世界」，情報処理学会創立 50 周年全国大会，http://staff.aist.go.jp/m.goto/IPSJ/event20100310.htm，(2010)．
3) 長嶋洋一：音楽的ビートが映像的ビートの知覚に及ぼす引き込み効果（1）---その第 1 報と実験計画---，http://1106.suac.net/news/docs/onchi-1.pdf，(2003)．
4) 長嶋洋一：音楽的ビートが映像的ビートの知覚に及ぼす引き込み効果（2）---心理学実験システムの開発とレイテンシの計測---, 情報処理学会研究報告, 2003-MUS-51, pp.83-90(2003)．
5) 岩宮眞一郎：音楽と映像のマルチモーダル・コミュニケーション，九州大学出版会（2000）．
6) 丸山健夫，安藤明人：音楽と映像のマッチング（1）---テンポと動き---，日本心理学会第 61

ⓒ 2011 Information Processing Society of Japan

回大会発表論文集，689（1996）.
7) 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-69，pp.31-36（2007）.
8) 飯塚太郎，Yue Yonghao，土橋宜典，西田友是：人間の知覚特性を考慮した音と映像の特徴検出および調和の許容時間を考慮したマッチング，情報処理学会研究報告，2008-AVM-63，pp.99-104（2008）.
9) バンダイナムコゲームス：THE IDOLM@STER OFFICIAL WEB，http://www.bandainamcogames.co.jp/cs/list/idolmaster.

ⓒ 2011 Information Processing Society of Japan