

音楽に合った舞踊動作の自動生成

二宮 啓^{†1} 中澤 篤志^{†1,†2} 竹村 治雄^{†1,†2}

近年コンピュータグラフィックスの分野では、リアルなキャラクタのアニメーションを生成する手法が数多く提案されてきている。我々は、人間の振る舞いを左右する感情を考慮した舞踊動作を対象として、入力した音楽信号から舞踊の表現に関する音楽情報を解析し、その結果に合った舞踊動作を生成する手法を提案してきた。動きデータからは動きのリズムと盛り上がりの特徴量を、音楽データからは楽曲構造解析によってセグメント分割し、特徴量としてリズム、盛り上がりを抽出する。舞踊動作生成時は、まず構造解析によって得られる音楽セグメント内のリズム成分と高い相関を示す動きの候補セグメントをすべて抽出する。そして最後に盛り上がり成分の相関を求めることで最適な動きセグメントを選択し、連結することで舞踊動作が生成する。またこの手法に対し、新たなモーションキャプチャデータを加えて動作生成し、生成された動作に対するブラインドユーザーテストによる有効性の評価、専門家による評価、ならびに一般ユーザーの利用に向けた WEB アプリケーションを開発した。

Dance to Music Character Animation

KEI NINOMIYA,^{†1} ATSUSHI NAKAZAWA^{†1,†2}
and HARUO TAKEMURA^{†1,†2}

This paper presents a new approach for synthesizing dance motion matched to input music based on the emotional aspects of dance performance. Our analysis method extracts motion rhythm and intensity from motion capture data and musical rhythm, structure, and intensity from musical signals. For synthesizing dance performance, we first find candidates of motion segment sets whose features are matched to those of music segments, and then we find the motion segment set whose intensity is matched to that of music segments. The experimental results indicate that our method actually creates dance performance as if the character was listening to the music.

1. はじめに

リアルなキャラクタアニメーションを生成する研究は、近年のコンピュータグラフィックスの分野で盛んである。これらの手法の大半がモーションキャプチャを利用しており、ユーザの入力したパス¹⁾、周囲の障害物²⁾、入力の音声情報³⁾やパートナーの動き⁴⁾などの入力情報をキーとして相関の取れた動きを生成する。このような研究の一つとして、我々は入力された音楽情報にうまく適合した舞踊動作を生成する手法を提案している⁵⁾。これは、人間が生まれながらにして持つ「音楽に合わせて舞踊を披露する」能力をコンピュータアニメーションの生成に応用したものである。つまり、ダンサーが行うアドリブダンス（未知の音楽に合わせて踊る）について考察すると、音楽に合わせて適当に体を動かしているのではなく、音楽を聴いて感じ取った情報や感情を基にして、それに合った動きのプリミティブを組み合わせることで舞踊を披露している。そこで我々は以下に示す仮定を設定し、アルゴリズムの設計を行った。

- 舞踊動作のリズムは音楽のリズムと同調する
- 舞踊動作の盛り上がりは音楽の盛り上がりと同調する

1つ目の仮定に関しては白鳥らの研究⁶⁾によって明かされている舞踊動作の特徴であり、音楽のリズムのタイミングと「留め動作」と呼ばれる舞踊動作中のキーポーズには高い相関があることがわかっている。2つ目の仮定に関しては舞踊動作に対する観察から得られた我々の知見であり、例えば「バラードのようなゆったりとした楽曲を聴いているときは気分的にもリラックスする」といった現象に由来している。この手法は動作解析、音楽解析、解析結果に基づいた動作生成の3つのステップから構成される。前処理として、舞踊動作のデータが収められたモーションキャプチャデータを複数用意しておき、動作解析によって得られた特徴量とともにデータベースに収めておく。入力は wav 形式の音楽音響信号で、音楽情景解析によって得られた特徴量と相関の取れる動きデータのセグメントをデータベースから選択し、つなぎ合わせることで新たな舞踊動作の生成を行う。動作解析では Laban によって提案された Weight Effort を用いて行う⁷⁾。すなわち、モーションキャプチャデータから Weight Effort を算出する方法を定義し、そこからリズム成分と盛り上がり成分

†1 大阪大学大学院 情報科学研究科

Graduate School of Information Science and Technology, Osaka University

†2 大阪大学 サイバーメディアセンター

Cybermedia Center, Osaka University

(以後、インテンシティ成分)を抽出する手法を用いた。一方音楽情景解析においては、まず ConstantQ Transform(CQT) と呼ばれる周波数成分算出方法⁸⁾を音楽音響信号に適用する。これにより、フーリエ変換より正確に各音符のスペクトルパワーを算出することが可能となる。この CQT 成分を用いてまず楽曲の構造解析を行い、音楽のセグメント分割を行う。また同時に CQT 成分からリズム成分とインテンシティ成分を算出し、音楽セグメントにこれらの特徴量を持たせる。動作生成においては、まずそれぞれの音楽セグメントに対して、リズム特徴量のみを用いて音楽セグメントと関連のとれる動きセグメントを抽出する。次に連続する動きセグメント間で、連結可能性を定量的な解析によってチェックし、自然な動きになると判断された動きセグメント間遷移のみを残す。そして最後に両者のインテンシティ成分に関してヒストグラムマッチングを行い、最適と評価された動きセグメント列を連結することで、舞踊動作を生成する。また、任意の音楽に対して本手法を用いて生成した動作と、生成の元となった動作を合わせたビデオを提示し、どちらの方がより音楽に合っているかを評価させる実験を行った。更に本手法を WEB 上で実行する環境を構築し、インターネット上で公開を行っている。

2. 動き特徴解析

動きの特徴量を求めるために Laban の提案した “Weight Effort”⁷⁾を用いる。Laban によると、人間の動きの印象を表す要素として、体の部位の動かし方に基づく “Effort” と、体の姿勢に基づく “Shape” の 2 つがある。Nakata らの行ったユーザスタディ⁹⁾から、Shape よりも Effort の方が動きの印象を左右することや、Effort の一種である Weight Effort と人間が感じる「動きの力強さ」との間に強い相関があることがわかっている。そこで、モーションキャプチャデータから Weight Effort $W(f)$ として人体の関節角度の線形和を算出し、そこからリズムと盛り上りを求める。

$$W(f) = \sum_{i=1}^N \alpha_i \cdot \dot{\theta}_i \quad (1)$$

ただし、 N は関節数、 $\dot{\theta}_i$ は各関節の回転速度、 α_i は重み係数である。この Weight Effort の定義と文献⁶⁾での解析結果を踏まえると、式(1)で計算される Weight Effort が極小となる部分が、音楽のリズムとの相関性の高い留め動作である可能性が高い。そこで、動きリズム特徴 F_R^{Motion} を式(2)のように定義し、キーフレームとする。このキーフレームは音楽のリズムと相関をとるために用いる。

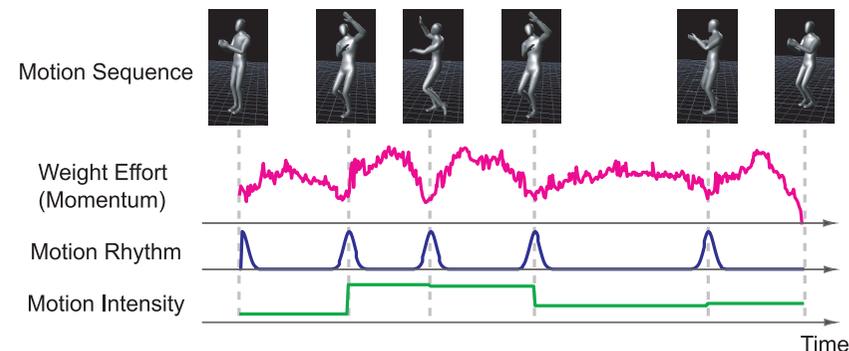


図 1 動きリズム特徴と動きインテンシティ(激しさ)特徴量

$$F_R^{\text{Motion}}(f) = \begin{cases} 1 & \text{if } W(f) \text{ is around the local minimum} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

一方、動きの強さに関しては、一定時間にどの程度関節が動作したか (Weight Effort が大きい) が重要となる。このため、連続する動きのキーフレーム間の平均をとることで、動きの大きさを表すこととする。これは、音楽の「盛り上がり」成分との相関評価に用いる。最終的に図 1 に示すような、動きリズム特徴と、動きインテンシティ(激しさ)特徴を得る。

3. 音楽特徴量解析

音楽を聴きながら舞踊動作を披露するときに重要な音楽特徴量として、我々は以下に示す 3 つの特徴量が特に重要であると考えた。

フレーズに基づく構造 ほとんどの楽曲は、フレーズと呼ばれるある一定の長さを持つ音符の流れの繰り返しによって構成されており、演奏家はこのフレーズを崩さないよう努めながら演奏を行う。これは舞踊においても同様である。そこで本手法では、繰り返しパターンに基づく構造解析を行い、音楽セグメントを抽出した。

リズム 楽曲のリズムは素人・玄人問わず最も認識しやすい音楽的要素の一つであることは明らかである。また、リズムに合わせて手を叩いたり身体を動かした経験を持つ人は非常に多く、舞踊にとっても非常に重要な要素であることがわかる。

ムード バラードのようなゆったりとした楽曲を聴いている時は落ち着いた気持ちになり、ロックのような激しい楽曲を聴いている時は興奮状態になることはしばしばある。この

ような現象は舞踊にとっても同様であり、さまざまな音楽要素が楽曲のムードを作り出しているが、本手法ではメロディラインの強さに着目した。

最初に入力の楽曲の音響信号に対してCQTを行い、音C3から音C6に渡る各音符のエネルギーを算出する(以後、CQTによって算出される時刻 t における音符 k のスペクトルパワーを $X(t, k)$ と表すこととする)。そして、音楽リズム成分と音楽インテンシティ成分を算出する一方、構造解析によって繰り返しフレーズに基づく音楽セグメントを求め、各々のセグメントに対して音楽リズム成分と音楽インテンシティ成分を持たせる。以下ではそれぞれの音楽特徴量の求め方について述べる。

3.1 繰り返しパターンに基づく楽曲構造解析

楽曲の構造解析を行う上で、以下の音楽的知識を参考にした。

[知識 1] 楽曲構造は数種類のフレーズの繰り返しによって構成される。

そこで、本ステップでは繰り返し構造を抽出し、その繰り返し構造に基づいてセグメント分割することを目的とする。構造解析において最も問題となることは、同じメロディやフレーズが異なる楽器によって違う音色で演奏される可能性があることである。楽器の音色の違いはその音の倍音構成に依存するため、様々な楽器で演奏される可能性のある楽曲の繰り返しパターンを認識するためには、倍音構成に対してロバストな方法を用いなければならない。

楽器の音色は、どの楽器であっても基音と倍音と呼ばれる基音の周波数の整数倍の周波数の音で構成される、という共通の周波数特性があり、音色の違いは基音や倍音のエネルギーの構成比率に由来する。そのため、周波数特徴量をそのまま用いるだけでは、この問題を克服することは困難である。

そこで、音色の違いにロバストな繰り返しパターンの抽出法¹⁰⁾をもとにして構造解析を行う。まず2つのCQT特徴量ベクトルの各要素の差分に対する自己相関関数を計算する。

$$r_{ij}(m) = \sum_{n=0}^{N-m-1} \Delta v_{ij}(n+m) \cdot \Delta v_{ij}(n) \quad (3)$$

ここで、 $\Delta v_{i,j}(n)$ は時間 i および j におけるCQT特徴量ベクトルの音符 n に対する差の絶対値を示す。

$$\Delta v_{ij}(n) = |X(i, n) - X(j, n)| \quad (4)$$

もし2つのCQT特徴量ベクトルが同じ音を含んでいる場合、前述の倍音構造の特徴より自己相関関数(式(3))はharmonic intervalと呼ばれる定まった間隔ごとにピークが現れ、異なる音が含まれている場合はピークの周期性は見られない。具体的には、自己相関関数 $r_{ij}(m)$ は主に $m = 0, 12, 19, 24, 29$ で強いピークが現れることから、これらに対しては少

ない重み付けを行い、それ以外の m に対して大きな重み付けを行うような重みベクトル W を設定し、自己相関ベクトル $R_{ij} = [r_{ij}(0), \dots, r_{ij}(N)]^T$ との内積を計算することで、2つのCQT特徴ベクトルの距離 d を求めることができる。

$$d_{ij} = W \cdot R_{ij} \quad (5)$$

この距離関数を用いて各時刻でのCQT特徴量ベクトルを計算し、その逆数を計算することで、類似度行列 S が得られ(図2(a))、以下に示す式でタイムラグ行列 T を算出する(図2(b))。

$$T_{ij} = S_{i,j+j} = \frac{1}{d_{i,i+j}} \quad (6)$$

タイムラグ行列に、多くの横線が現れているのが分かる。そこで、画像処理でよく用いられるErosionとDilationの処理を行うことで横線を明確にし(図2(c))、しきい値処理を施すことで繰り返しパターンを抽出する(図2(d))。この繰り返しパターンに基づいてセグメント分割をすることで、音楽セグメントを抽出する。

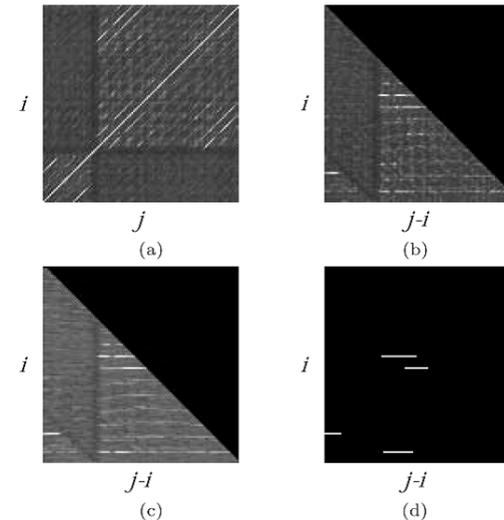


図2 楽曲の構造解析における各過程での結果。(a)類似度行列 S_{ij} 、(b)タイムラグ行列 $T_{i,j-i}$ 、(c)ErosionとDilationの処理後のタイムラグ行列、(d)繰り返しパターン抽出の結果(文献⁵⁾より引用)

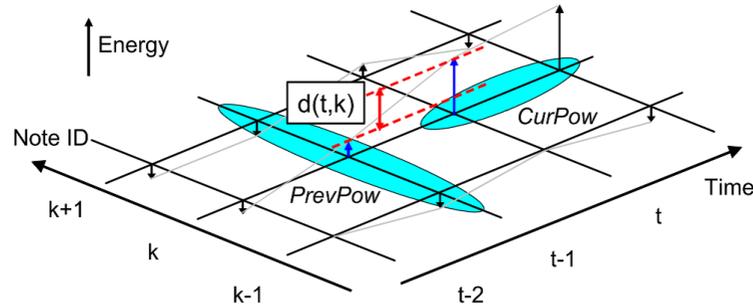


図3 発音成分の算出 (文献⁵⁾ より引用)

3.2 リズムトラッキング

リズム時刻を推定する上では、以下の音楽的知識を用いる。

[知識 2] 発音はリズム時刻で行われる可能性が高い。

[知識 3] 発音成分のピークの間隔はリズム間隔とほぼ一致する。

まず、知識 2 を用い、時刻 t 、音 k における発音の強さ $d(t, k)$ を式 (7) により算出する¹¹⁾。

また図 3 に発音成分算出の概要を示す。

$$d(t, k) = \begin{cases} \max(X(t, k), X(t+1, k)) - \text{PrevPow} \\ \quad : \text{if } \min(X(t, k), X(t+1, k)) \geq \text{PrevPow} \\ 0 \quad \text{otherwise} \end{cases} \quad (7)$$

$$\text{PrevPow} = \max(X(t-1, k), X(t-1, k \pm 1))$$

全音符にわたる発音成分の合計 $D(t) = \sum_f d(t, f)$ を求める。ここから、時刻 t における k 番目の音符の発音の強さの算出が可能となる。

次に知識 3 に基づき、発音成分 $D(t)$ の自己相関関数を求めることによって平均リズム間隔を算出し、パルス間隔が平均リズム間隔となっているパルス列と発音成分 $D(t)$ との相互相関関数を求めることによって、リズムの開始時刻を算出する。しかし、実際の楽曲ではリズム間隔が常に一定ではなくわずかながら変化しており、この誤差が累積するとリズムの追従が不可能になる。そこで、再び知識 2 に基づいて、求めたリズム時刻の周辺で発音成分 $D(t)$ のローカルピークを抽出することによって、より正確なリズム時刻の推定を行う。この結果を用いて、音楽セグメント M に対する音楽リズム成分 F_R^{Music} は以下のように得ら

れる。

$$F_R^{\text{Music}}(f; M) = \begin{cases} 1 & \text{if } f \text{ in } M \text{ is rhythm time} \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

3.3 音楽インテンシティ成分の算出

楽曲の盛り上がり解析を行う上で、メロディラインのエネルギーの大きさに着目する。メロディラインに関しては、以下の音楽的知識が分かっている。

[知識 4] 楽曲が盛り上がると、メロディラインのパワーが増加する。

[知識 5] メロディラインは C4 の音より高い音域で演奏される。

最初に構造解析で求めた音楽セグメント M における CQT 特徴量ベクトルの時間平均 \bar{X} を求める。次に求めた平均ベクトルの周波数方向の極大値のみを抽出した平均音素ベクトル X_{peak} を抽出する。

$$X_{\text{peak}}(M, k) = \begin{cases} \bar{X}(M, k) & \text{if } \bar{X}(M, k) > \bar{X}(M, k \pm 1) \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

この平均音素ベクトルのうち、音 C4 から C6 の要素の音圧レベルを抽出することによって、音楽インテンシティ特徴量とした。

4. 動きおよび音楽特徴量を考慮した舞踊動作生成

得られた動き特徴量と音楽特徴量に基づいて、舞踊動作を新たに生成する。図 4 にその概略を示す。まずリズム特徴量のマッチングによって、動きセグメントを抽出する。次にすべての動きセグメント対に対して連結可能性を調べ、連結したときに自然な動きとなる対のみを候補として残す。そして最後にインテンシティ特徴量のマッチングを行い、最も関連の高いものを出力とする。

4.1 リズム特徴量に基づいた動きセグメントの抽出

まず、入力音楽の構造解析で得られた音楽セグメントにおける音楽リズム特徴量と高い相関を示す動きリズム特徴量を持つような動きセグメントを、動きデータベースから抽出する。今、長さ L_{music} の音楽セグメント M に対応する動きセグメントの抽出を行う場合を考える。動きデータベースの中の解析対象とする動きデータの長さを L_{motion} 、解析対象となる動きの部分の開始フレームを f_0 とする。また、解析時にはスケールパラメータ $s \in [0.9, 1.1]$ を用いて動きの伸縮を可能にしている。このときのリズム特徴量の相関値は式 (10) で求め

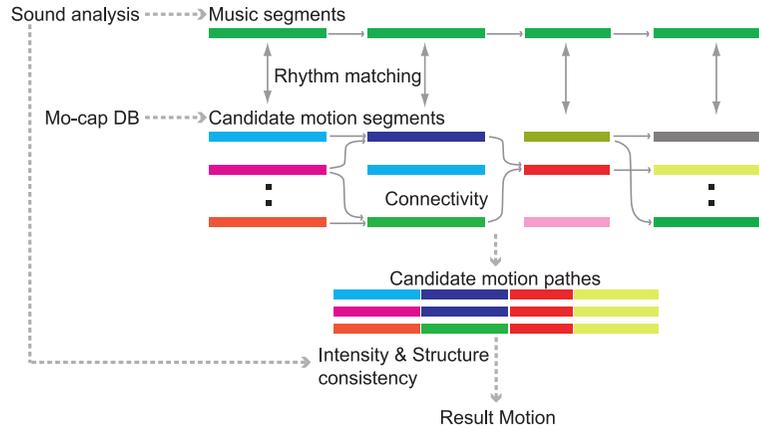


図4 舞踊動作生成アルゴリズム

る。閾値以上のものを動きセグメントの候補として保存しておく。この処理をすべての音楽セグメントに対して、対象とする動きデータ、 f_0 の値、スケールパラメータを変更しながら行うことで、各々の音楽セグメントに対応する動きセグメント群を求める。

$$\arg \max_s \sum_{f=0}^{L_{\text{music}}} \frac{F_R^{\text{Music}}(f; M) \cdot F_R^{\text{Motion}}(s \cdot f + f_0)}{F_R^{\text{Music}}(f; M) + F_R^{\text{Motion}}(s \cdot f + f_0)} \quad (10)$$

$$f_0 \in [0, L_{\text{motion}} - L_{\text{music}}]$$

4.2 動きセグメントの連結可能性

現段階で得られている動きセグメントはリズムが反映されているが、実際に舞踊動作を生成した際に自然な動きになっているか、については保証されていない。そこで、2つの動きセグメントの連結に関する評価関数を設け、その評価値に対して閾値処理を行うことで、連結しても自然な動作であるかを判断する。この連結可能性の評価に対しては、連結元セグメントの最終フレームと、連結先セグメントの先頭フレームの間で、姿勢の相関と動きの相関を評価することで行う。両者の違いが閾値以下であった場合、連結可能であると評価する。

4.3 インテンシティ成分の類似度計算

次に、複数ある動きセグメント列CSの中から、インテンシティ成分の類似度計算により最も相関の取れた動きセグメント列を一つ選び出す。ここで、音楽セグメント列および動きセグメント列のインテンシティ成分をヒストグラムとみなし、バタチャリア係数を算出する

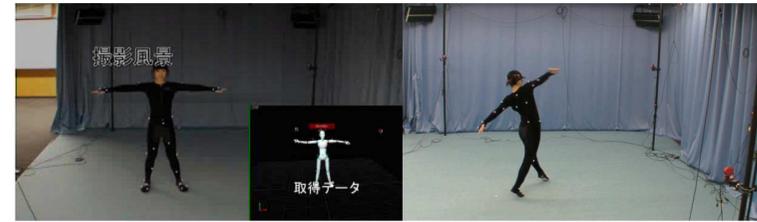


図5 実験データの取得シーン

ことで、類似度を算出する。バタチャリア係数を最大化する動きセグメント列を取り出すことによって、音楽の特徴量と最もよく相関しており、かつ連結しても不自然にならない動きセグメント列を作り出すことができる。

4.4 連結動作の生成

最後に、選ばれた動きセグメント列のセグメント間を連結動作でつなげることで、舞踊動作の生成が完了となる。姿勢に関しては、リンクの方向ベクトルを1次および2次微分の連続性を考慮しつつ、スプライン補間を行う。またキャラクタの位置に関しては、キャラクタの動作中に足が地面に対してすべる、というCG特有の問題を考慮しながら補間を行う。

5. 実験

手法の有効性を確かめるために、新たに動作データを取得して動作生成を行った。また被験者に対して、本手法で生成した生成動作データと他の動作データを被験者にランダムに提示し、音楽に合っている動作か否かを評価尺度で評価させるブラインドテストを行い、手法の有効性を確かめた。また、生成した舞踊を取得した舞踊データの踊り手に見せ、意見を求めた。データ取得には、OptiTrack社の光学式モーションキャプチャシステム(14台の赤外線カメラ(FLEX:V100)で構成)を用いて舞踊動作の撮影スタジオを構成し、100Hz、34マーカーポイントでデータを取得した。

5.1 実験データ

実験データとして、神戸女学院大学音楽学部音楽学科舞踊専攻の学生9人が演じた舞踊のデータを取得した(図5)。それぞれ3~5曲のテンポや雰囲気の違いのある音楽に合わせて、踊り手の専門とする種類の舞踊を踊った。踊り手の専門とする舞踊は、主にモダンダンス4名(計84ファイル、約37分)、バレエ2名(計36ファイル、約16分)、朝鮮舞踊1名(11ファイル、約5分)、コンテンポラリーダンス2名(計30ファイル、約11分)である。

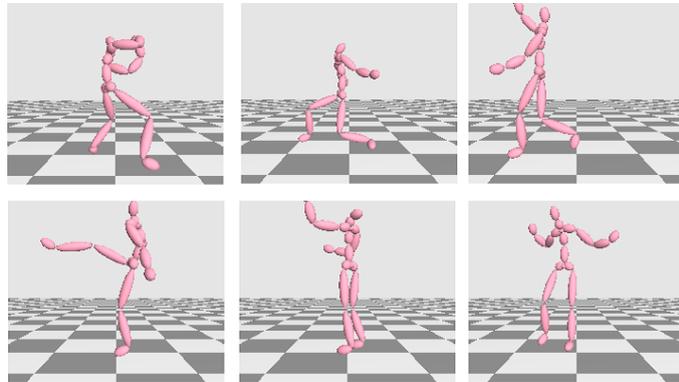


図 6 生成した舞踊動作の例

表 1 評価実験の結果 (評価得点)

音楽	無作為に選んだ	提案手法で生成
	舞踊	した舞踊
Thousand Dreams	3.0	3.6
HANABI	3.3	3.8
Jerk It Out	3.1	3.9
Viva la Vida	2.5	3.7
ネイティブダンサー	3.6	3.9
平均	3.1	3.78

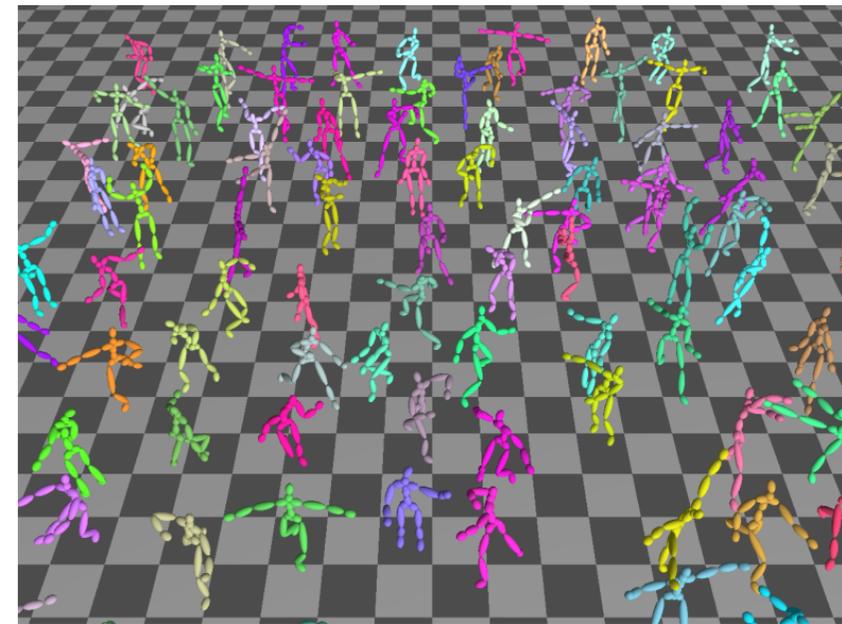


図 7 同一音楽に対して 100 種類の動作を生成した例

5.2 動作生成実験

取得したデータの中から 1 人分 (表 2 の舞踊者 1) のデータを用いて CD から取り込んだ wav 形式の音楽を入力として、提案手法で舞踊動作を生成した。ここで、取得した全データを用いずに 1 人分のデータだけを用いたのは、複数の踊り手が踊ったデータを混在させて舞踊動作の生成を行うと、異なる踊り手の特徴が混ざり違和感が生じると考えられるためである。図 6 は生成された舞踊動作の例である。また、図 7 は 1 つの音楽に対して 100 個の異なる舞踊動作を生成し、同時に踊らせた例である。

5.3 生成動作に対するユーザー評価実験

本手法により生成した舞踊のリズムが生成に用いた音楽のリズムと合っていることを確かめ、提案手法の有用性を示すために評価実験を行った。前述した舞踊データを用いて、30 秒程度の 5 曲の音楽に対して 1 つずつ舞踊動作を生成し、ビデオを生成した。また、提案

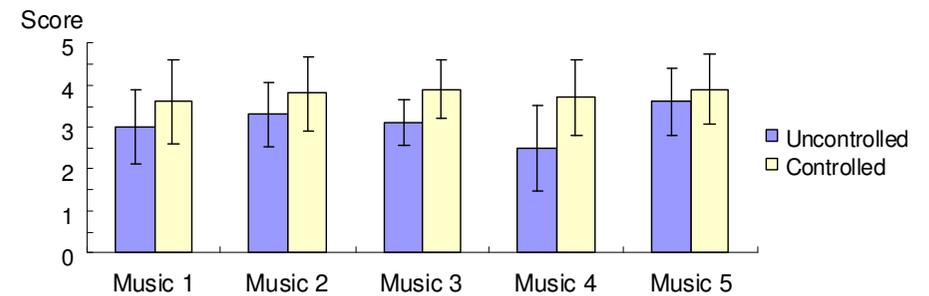


図 8 評価実験の結果 (平均評価得点と標準偏差)

手法によって生成した舞踊との対照実験を行うため、舞踊生成に用いた元の舞踊データの中から無作為に選んだ舞踊動作を音楽に合わせたビデオを用意した。これら 10 個の舞踊動作をランダムな順番で 10 人の評価者に見せ、「音楽のリズムと舞踊動作のリズムが合っているか」を 5 段階で評価させた。また、この評価実験を行った後に評価者に本手法を用いて生成した舞踊だけを見せ、主観評価をもとめた。

評価実験の結果を表 1, 図 8 に示す。また、評価者に提案手法によって生成した舞踊だけを見せ、意見を求めた結果、以下のような指摘があった。

[指摘 1] 舞踊動作中に動きがおかしくなることがある

[指摘 2] 音楽の雰囲気と動きの雰囲気が合っていない場合がある

5.4 踊り手による評価

生成したいいくつかの舞踊を舞踊の生成に用いたデータの踊り手に見せて意見を求めたところ、以下のような指摘があった。

[指摘 A] 音楽のリズムに合っている

[指摘 B] 舞踊にメリハリがない

[指摘 C] ためるような動作の後にはじける動きがないことがある

[指摘 D] 自分が思いつかないような動きのつなぎ方をするので動きのつながりを見ていておもしろい

[指摘 E] 動きのつなぎ目がおかしい場合がある

6. 考 察

6.1 提案手法の有効性

5.1 節の図 6 で示すようにオープンなデータベース (CMU Motion Capture Database) を用いた白鳥らの研究に比べ、新たに用意した舞踊動作データを用いた本結果は、明らかに優れている結果を得ることができた。5.1 節の図 7 で示すように、本手法では、セグメント間の遷移に対して様々なケースをとることが可能なため、同一の音楽に対して無限のパターンの舞踊動作を生成することができる。これはダンサーが行う舞踊が、人ごとに異なったり、状況に応じて異なったりすることとよく類似する。さらに、指摘 D にあるように、人が思いつかないような動きのつなぎ方をした動作が生成できるという非常に興味深い意見も得られている。

また、5.2 節の結果から 5 曲全てにおいて提案手法による舞踊動作の方がリズムに合っているという評価を得、本手法により生成した舞踊のリズムが生成に用いた音楽のリズム

と合っていると見える。また、提案手法により生成した結果と無作為に選んだ結果に対する評価の対を標本として、両者の結果に差がないと仮定して分散分析を行った結果 p 値は $p = 0.00026 (\ll 0.01)$ となり明らかに有意な水準で棄却されることから、統計的にも有意に本手法が音楽に合った動作データを生成することができるといえる。さらに、指摘 A にあるように踊り手から見ても音楽のリズムに合っているという意見が得られている。

6.2 提案手法の問題点

6.2 節で挙げた有効性がある一方で、提案手法にはいくつかの問題点がある。まず、指摘 1 および指摘 E にあるように舞踊の途中で動きがおかしくなることがあることである。生成に用いている元のデータ問題がないことや指摘 E から、動きのおかしさは動きのつなぎ目部分で生じていると考えられる。これは、動きセグメントの候補間の接続可能性の評価関数や補間方法に問題があるためであると考えられる。接続可能性の評価関数については、評価関数中の閾値や関節ごとの重みなどのパラメータの調整、評価に用いる物理量の再検討 (例えば角運動量を用いる) などの改良方法が考えられる。補間方法については、現在足がすべる問題の対処をあまり重点的に行っていないので、両足の位置関係や地面に足が着いているかによって場合分けをし、それぞれの場合によって接続不可能と判定したり、足の位置を修正することで大幅に改善されることが考えられる。

次に、指摘 2 にあるように音楽の雰囲気と舞踊の雰囲気が合っていない場合があることである。この原因としてまず、舞踊の種類と曲の種類が合っていないことが挙げられる。例えば、ヒップホップの曲に対してバレエの動きを使って舞踊を生成した場合などである。これは、様々な種類の舞踊データを取得し、舞踊の種類ごとに動きデータベースを分け、音楽の種類に合った舞踊を選択することで解決できる。次に、インテンシティ特徴量のマッチングがうまくいっていないことやインテンシティ特徴量のマッチングだけでは、音楽と舞踊の雰囲気が合わせられないことが考えられる。これを確かめるためにまず、インテンシティ特徴量のマッチングに有効性があるかどうかの評価実験を行う必要がある。その上で、インテンシティ特徴量のマッチング方法の変更や、他の特徴量の導入などの対応が必要である。

最後に、指摘 B にあるように生成した舞踊にダンサーが踊る舞踊のようなメリハリがないことが挙げられる。この原因として、まず、インテンシティ特徴量のマッチングがうまくいっていない可能性があることが挙げられる。これについては、前述のように評価実験を行ったあとで対応する必要がある。次に、指摘 C にあるように、ためる動きとはじける動きを考慮する必要があることが挙げられる。音楽の激しさの立ち上がり (急激に激しくなる) の前後にためる動きとはじける動きをあてるようにできれば、メリハリのある舞踊

を生成することができると考えられる。このために、Weight Effort を用いて舞踊データから動きが急激に激しくなる部分を見つけ、音楽のインテンシティが急激に上がるような方法が考えられる。

7. WEB アプリケーション

本手法を WEB アプリケーションとして実装し、インターネット上で一般公開している¹²⁾。WEB フォーム上で wav ファイルを送信することで、その音楽からビートおよび盛り上がり自動的抽出し、その音楽にあった動作の動作データ (BVH ファイル) としてダウンロードできる。同サイトで配布しているビューアーを用いることで、元の音楽データと動作データを同期した形で閲覧可能であり、また BVH をインポートできるソフトウェアで、音楽にあったキャラクターアニメーションを生成することができる。

8. 結論と今後の課題

本論文では、我々が提案したモーションキャプチャデータを用いて「音楽に合った」新規動作データ (舞踊動作) を生成する手法について述べ、新たに動作データを取得・生成実験を行うことでより自然な動作を生成することを可能にした。また、生成した動作データと音楽とをユーザーに提示するブラインドテストを行い、本手法で生成した結果が明らかに有意に音楽に合った舞踊動作アニメーションを生成できることを確認した。本手法では、セグメント間の遷移に対して様々なケースをとることが可能なため、同一の音楽に対して無限のバリエーションの舞踊動作を生成することができる。また本手法は WEB アプリケーションとして公開しており、WEB フォーム上でサウンドファイルを送信することで、その音楽からビートおよび盛り上がり自動的抽出し、その音楽にあった動作の動作データ (BVH ファイル) としてダウンロードできる。

今後の課題として、6.2 節で挙げたようにつなぎ目が自然になるように手法を改良することや、生成した舞踊の盛り上がり音楽の盛り上がりと合っているかの評価実験を行うこと、メリハリのある舞踊を生成できるように手法を拡張することが挙げられる。また、さらなる課題を見つけるために、音楽に合わせて踊り手が踊った舞踊と、本手法を用いて生成した舞踊との比較実験を行うことも挙げられる。さらに、本手法の応用としてユーザコントロールを取り入れて生成する舞踊をユーザの好みに合うように調整できるようにすることや、処理をリアルタイム化して、音楽に合わせた舞踊をリアルタイムで生成するアプリケーションを実装することが考えられる。

謝辞 本研究は、総務省 SCOPE 若手 ICT 研究者育成型研究開発 (ICT 安心・安全技術) 「三次元データ認識による災害状況自動計測システムの研究開発」の支援を受けて行われた。また、本研究は文部科学省科学研究費「遮蔽や人体形状変化に頑健な非装着型モーションキャプチャ」の補助を受けて行われた。

参考文献

- 1) L. Kovar, M. Gleicher and F. Pighin: "Motion graphs", ACM Trans. on Graphics, 21, 3, pp. 473-482, 2002.
- 2) J. Lee, J. Chai, P. S. A. Reitsma, J. K. Hodgins and N. S. Pollard: "Interactive control of avatars animated with human motion data", ACM Trans. on Graphics, 21, 3, pp. 491-500, 2002.
- 3) M. Stone, D. DeCarlo, I. Oh, C. Rodriguez, A. Stere, A. Lees and C. Bregler: "Speaking with hands: Creating animated conversational characters from recordings of human performance", ACM Trans. on Graphics, 23, 3, pp. 506-513, 2004.
- 4) E. Hsu, S. Gentry and J. Popović: "Example-based control of human motion", Proc. SIGGRAPH/Eurographics Symposium on Computer Animation 2004, pp. 69-77, 2004.
- 5) Takaaki Shiratori, Atsushi Nakazawa, Katsushi Ikeuchi, "Dance-to-Music Character Animation," Computer Graphics Forum (EUROGRAPHICS2006), Vol. 25, No. 3, pp.449-458, 2006.
- 6) 白鳥貴亮, 中澤篤志, 池内克史, "モーションキャプチャと音楽情報を用いた舞踊動作解析手法," 電子情報通信学会論文誌, Vol.J88-D2 No.8 pp.1662-1671, 2005.
- 7) R. Laban and L. Ullmann: "Mastery of Movement", Princeton Book Company Publishers, 1960.
- 8) J. C. Brown: "Calculation of a constant Q spectral transform", Journal of Acoustic Society of America, 89, 1, pp.425-434, 1990.
- 9) Nakata T and Mori T and Sato T: "CEntertainment and Amusement Robot Technologies. Analysis of Impression of Robot Bodily Expression", J Rob Mechatron, 14, 1, pp.27-36, 2002.
- 10) M. Wang and L. Lu and H.J. Zhang: "Repeating pattern discovery from acoustic musical signals", Proc. IEEE Int'l Conf. on Multimedia and Expo, pp.2019-2022, 2004.
- 11) M. Goto: "An audio-based real-time beat tracking system for music with or without drum-sounds", Journal of New Music Research, 30, 2, pp.159-171, 2001.
- 12) "The Dance to Music Character Animation Project", <http://d2m.ime.cmc.osaka-u.ac.jp/>, 2009.