

音楽情景を考慮した舞踊動作生成手法

白鳥 貴亮[†] 中澤 篤志^{††} 池内 克史[†]

[†] 東京大学 生産技術研究所 〒153-8505 東京都目黒区駒場 4-6-1

^{††} 大阪大学 サイバーメディアセンター 〒560-0043 大阪府豊中市待兼山町 1-32

E-mail: [†]{siratori,ki}@cvl.iis.u-tokyo.ac.jp, ^{††}nakazawa@cmc.osaka-u.ac.jp

あらまし 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ー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。しかし、人間の振る舞いを左右する感情を考慮した手法はほとんど提案されていない。そこで本稿では、感情が大きな要因となる舞踊動作を対象として、入力音楽信号から舞踊の表現に関する音楽情景を解析し、その結果に合った舞踊動作を生成する手法を提案する。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される。さまざまな楽曲や動きデータに適用して実験を行い、その有効性を示す。

キーワード モーションキャプチャ、音楽情報、動き特徴解析、音楽情景解析

Character Animation for Dancing to Music

Takaaki SHIRATORI[†], Atsushi NAKAZAWA^{††}, and Katsushi IKEUCHI[†]

[†] Institute of Industrial Science, The University of Tokyo Komaba 4-6-1, Meguro-ku, Tokyo, 153-8505 Japan

^{††} Cybermedia Center, Osaka University Machikaneyama 1-32, Toyonaka, Osaka 560-0043 Japan

E-mail: [†]{siratori,ki}@cvl.iis.u-tokyo.ac.jp, ^{††}nakazawa@cmc.osaka-u.ac.jp

Abstract This paper presents a new approach for synthesizing dance performance 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 a character was listening to the music.

Key words Motion Capture, Musical Information, Motion Feature Analysis, Auditory Scene Analysis

1. はじめに

リアルなキャラクター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は重要なトピックになりつつある。その既存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。

本稿で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する。人間は、生まれながらにして「音楽に合わせて舞踊を披露する」という能力を持ち合わせており、誰しも一度はリズムカルな音楽に合わせて体を動かしたことがあると思われる。踊り手がアドリブダンスを披露する際につい

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。そこで本研究で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

一つ目の仮定に関しては、Shiratori らの研究 [5] 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。二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。

本手法は主に動作解析、音楽解析、解析結果に基づいた動作生成の3つのステップから成り立つ。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 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。動作解析では主に Laban によって提案された “Weight Effort” を用いて行う [6]。Laban によると、人間の動きの印象を左右する要素として、力強さなど身体の部位の動かし方に基づく “Effort” と身体の姿勢に基づく “Shape” の2つに大きく分けられるとしている。また Nakata らの行ったユーザスタディ [7] によると、Shape より Effort の方が動きの印象を左右する、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、などがわかっている。そこで、我々は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の成分 (以後、インテンシティ成分) を抽出する手法を提案する。一方音楽情景解析においては、まず Constant Q 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相関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。

以降では、2章にて動きの特徴量解析について、3章にて音楽の特徴量解析について述べる。次に4章ではこれらの解析結果を基にした舞踊動作の生成について述べる。5章で実験結果について述べ、最後に6章で本稿をまとめる。

2. 動き特徴量解析

我々の動作解析においては、Laban の提案した “Effort” という動きの要素のうち、“Weight Effort” が非常に重要な役割を担っている。本章では、Weight Effort を定量的に定義し、動き特徴量を抽出する方法について述べる。

2.1 人体モデル

本手法で用いた人体モデルを図1に示す。人体姿勢は体中心座標系と呼ばれる座標系に各フレームごとに変換される。体中心座標系では腰の中心を原点とし、体の横方向を x 軸、正面方向を y 軸、体の鉛直方向を z 軸とし、それぞれを単位ベクトル r_x , r_y , r_z で表す。世界座標系での原点の軌跡は $t = (t_x, t_y, t_z)^T$

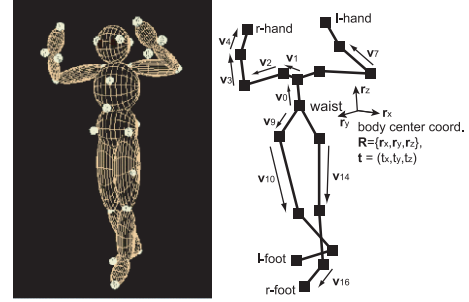


図1 人体モデル

で表す。また各リンクの方向ベクトルは、体中心座標系における単位ベクトルとして計算され、 v_n (n は各リンクに割り当てられたインデックス) で表される。

2.2 Weight Effort の算出と動き特徴量の抽出

Laban の定義によると、Weight Effort は動きの強さを表す。そこで本手法では、フレーム f に対する Weight Effort $W(f)$ を、各リンクに対する重み α とすべてのリンクのモーメントと体の向きの変化量の線形和によって計算する。

$$W(f) = \sum_{i=0}^{16} \alpha_i \arccos\left(\frac{\dot{v}_i(f)}{|\dot{v}_i(f)|} \cdot \frac{\dot{v}_i(f+1)}{|\dot{v}_i(f+1)|}\right) + \sum_{j \in \{x, y, z\}} \arccos\left(\frac{\dot{r}_j(f)}{|\dot{r}_j(f)|} \cdot \frac{\dot{r}_j(f+1)}{|\dot{r}_j(f+1)|}\right) \quad (1)$$

この Weight Effort の定義と Shiratori らの解析結果 [5] をふまえると、式1で算出される Weight Effort が極小となる部分で舞踊のキーポーズである「留め動作」である可能性が高いということになる。そこで、動きリズム成分 F_R^{Motion} は式2のように定義し、この極小点をキーフレームと呼ぶことにする。

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

一方 Nakata らの研究 [7] によれば、Weight Effort と動きの盛り上がりに対する印象には強い相関があることがわかっている。そこで、本手法では Weight Effort に加え全身の正面方向の速度 $r_y \cdot \dot{t}$ も考慮に入れ、まず各フレームごとに動きインテンシティ成分 $I(f)$ を算出する。

$$I(f) = W(f) \cdot (1.0 + k \cdot r_y(f) \cdot \dot{t}(f)), \quad (3)$$

そして連続する動きのキーフレーム間 $[f_i^R, f_{i+1}^R]$ で平均を取ること、動きのインテンシティ成分 $F_I^{\text{Motion}}(f)$ ($f \in [f_i^R, f_{i+1}^R]$) とした。

$$F_I^{\text{Motion}}(f) = \sum_{f=f_i^R}^{f_{i+1}^R} I(f) / (f_{i+1}^R - f_i^R + 1) \quad (4)$$

図2に、抽出された動き特徴量の例を示す。

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。

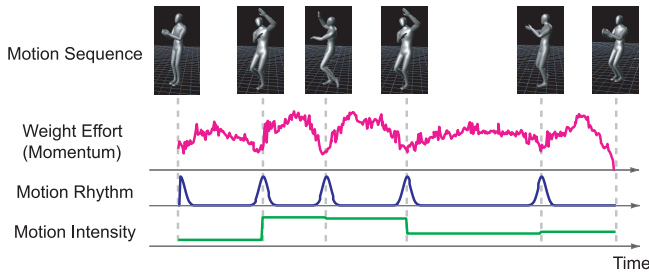


図 2 動き特徴量の例

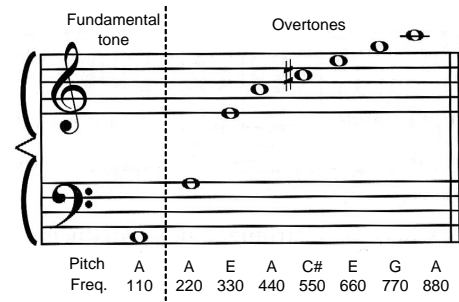


図 3 音 'A'(110Hz) を出したときの周波数特性

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 パ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。

そこで本過程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。構造解析においてもっと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演奏される可能性があることである。楽器の音色はその周波数特性によって決まるため、周波数領域で繰り返しパターンを認識するためには、楽器の音色に対してロバストな方法でなければならない。実際、人間の聴覚では異なる楽器で演奏されたフレーズでも同一の音符列だと認識することが可能である。

図 3 に音色の周波数特性の例を示す。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違いに由来する。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類似度の算出方法 [9] を基にして、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。

$$r_{ij}(m) = \sum_{n=0}^{N-m-1} \Delta v_{ij}(n+m) \cdot \Delta v_{ij}(n) \quad (5)$$

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

$$\Delta v_{ij}(n) = |X(i, n) - X(j, n)| \quad (6)$$

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 (5) は *harmonic interval* と呼ばれる定まった間隔毎にピークが現れ、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 \mathbf{W} を設定し、自己相関ベクトル $\mathbf{R}_{ij} = [r_{ij}(0), \dots, r_{ij}(N)]^T$ との内積を計算することで、2 つの CQT 特徴量ベクトルの距離 d を求めることができる。

$$d_{ij} = \mathbf{W} \cdot \mathbf{R}_{ij} \quad (7)$$

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

$$T_{ij} = S_{i,i+j} = \frac{1}{d_{i,i+j}} \quad (8)$$

図 4 にそれぞれの行列の例を示す。タイムラグ行列に、多くの横線が現れているのがわかる。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、閾値処理によって抽出することで繰り返しパターンが抽出できる。この繰り返しパターンに基づいてセグメント分割することで、音楽セグメントを抽出する。

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 音の発音時刻とリズム時刻は一致する可能性が高い

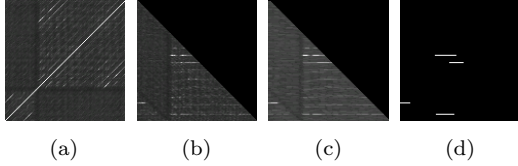


図 4 楽曲の構造解析における各過程での結果．(a) 類似度行列，(b) タイムラグ行列，(c) Erosion と Dilation の処理後のタイムラグ行列，(d) 繰り返しパターン抽出の結果

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 9 により算出する [10]．また図 5 に発音成分算出の概要を示す．

$$d(t, k) = \begin{cases} \max(X(t, k), X(t+1, k)) - \text{PrevPow} \\ (\min(X(t, k), X(t+1, k)) \geq \text{PrevPow}) \\ 0 \quad (\text{otherwise}) \end{cases} \quad (9)$$

$$\text{PrevPow} = \max(X(t-1, k), X(t-1, k \pm 1)) \quad (10)$$

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 において 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 \mathcal{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．

$$F_R^{\text{Music}}(f; \mathcal{M}) \begin{cases} 1 & \text{if } f \text{ in } \mathcal{M} \text{ is estimated rhythm time} \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

3.3 音楽インテンシティ成分の算出

楽曲のムード解析を行う上で、本手法ではメロディラインのエネルギーの大きさに着目した．メロディラインに関しては、以下にあげる音楽的知識がわかっている．

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本手法では、最初に構造解析で求めた音楽セグメント \mathcal{M} における CQT 特徴ベクトル \bar{X} の時間平均を求める．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．

$$X_{\text{peak}}(\mathcal{M}, k) = \begin{cases} \bar{X}(\mathcal{M}, k) & \text{if } \bar{X}(\mathcal{M}, k) > \bar{X}(\mathcal{M}, k \pm 1) \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

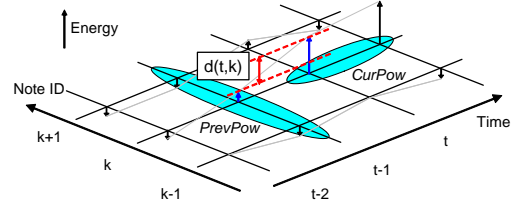


図 5 発音成分の算出

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 F_I^{Music} とした．

$$F_I^{\text{Music}}(f; \mathcal{M}) = \log_{10} \left(\sum_{k \in [C4, C6]} X_{\text{peak}}(\mathcal{M}, k)^2 \cdot f_k^2 \right) \quad (13)$$

ただし、 f_k は音 k の周波数を示す．

4. 動きおよび音楽特徴量を考慮した舞踊動作生成

2 章によって得られた動き特徴量とおよび 3 章によって得られた音楽特徴量に基づいて、舞踊動作を新たに生成する．図 6 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、最も相関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

「舞踊動作は音楽のリズムとマッチしていなければならない」という仮定に基づき、始めに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データベースから抽出する．より詳細な説明をするために、ここでは長さ L_{motion} をもつ動き列と、長さ L_{music} の長さを持つ音楽セグメント \mathcal{M} について考える．本手法ではスケールパラメタ $s \in [0.9, 1.1]$ を用いて、動きデータの伸縮を可能にしている．動きセグメントの開始フレームを f_0 として、式 (14) を用いて音楽セグメントのリズム特徴量との相関値が最も高くなるスケールパラメタとそのときの相関値を求める．

$$\arg \max_s \sum_{f=0}^{L_{\text{music}}} \frac{F_R^{\text{Music}}(f; \mathcal{M}) \cdot F_R^{\text{Motion}}(s \cdot f + f_0)}{F_R^{\text{Music}}(f; \mathcal{M}) + F_R^{\text{Motion}}(s \cdot f + f_0)} \quad (14)$$

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

求めた相関値に対し閾値処理をすることで、抽出された動きセグメントの絞込みを行う．この処理をすべての音楽セグメントに対して行うことで、各々の音楽セグメントとリズム特徴量が類似した動きセグメント群を求める．

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、についてはまったく保証されていない．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

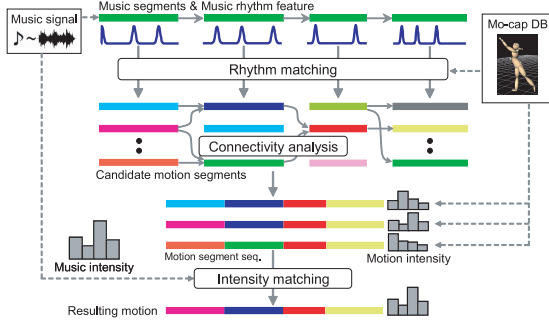


図 6 舞踊動作生成のながれ

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、を判断した。

この評価関数は、キャラクタの姿勢に関する類似度関数評価関数 S_{pose} と連結部分前後での動きに関する類似度評価関数 S_{move} の 2 つによって構成される。いま、動きセグメント A のフレーム i^A と動きセグメント B のフレーム j^B との間でこれらの評価値を求めることを考える。姿勢の類似度評価関数 S_{pose} は、本手法で用いている人体モデルでの各リンクの方向ベクトル \mathbf{v} の内積と各リンクに対する重み β により計算される。

$$S_{\text{pose}}(i^A, j^B) = \sum_l \beta_l (\mathbf{v}_l(i^A) \cdot \mathbf{v}_l(j^B)) \quad (15)$$

一方動きの類似度評価関数 S_{move} については、リンクの方向ベクトルの時間微分、すなわち速度ベクトルにより計算される。この速度ベクトルは、関数 h を通して同次座標系における単位ベクトルに変換され、その内積を算出することでベクトルの大きさと向きを同時に評価することが可能となる。連結前のオリジナルの動き列の中でのリンクの方向ベクトルの速度 $\dot{\mathbf{v}}$ と、連結後の大まかなリンクの方向ベクトルの速度 $(\mathbf{v}_l(j^N) - \mathbf{v}_l(i^M))$ との内積で比較することにより、動きの類似度の算出が可能となる。

$$S_{\text{move}}(i^A, j^B) = \prod_l g[h((\mathbf{v}_l(j^B) - \mathbf{v}_l(i^A))) \cdot h(\dot{\mathbf{v}}_l(i^A))] \cdot g[h((\mathbf{v}_l(j^B) - \mathbf{v}_l(i^A))) \cdot h(\dot{\mathbf{v}}_l(j^B))] \quad (16)$$

ただし、

$$g[x] = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

である。

最終的に連結可能性は、4.1 節で得られた動きセグメント群のうち、時間的に連続する 2 つの動きセグメント群のすべての組み合わせで、評価値 $S_{\text{pose}} + S_{\text{move}}$ を算出し、閾値処理をする。これにより、リズムに関して相関が取れており、かつ連結しても自然に見える複数の動きセグメント列を抽出することが可能となる。

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

表 1 リズムトラッキングの結果

曲名 (ジャンル)	リズム [sec] ([bpm])
<i>T-Connection</i> (break)	0.536 (112)
<i>La Cumparsita</i> (tango)	0.454 (132)
<i>Carmen Suite</i> (classic)	0.417 (144)
<i>Nutcracker Suite</i> (classic)	0.714 (84)

列の一つを選び出す。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 [11] を算出することで、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。

$$\arg \max_{D \in CS} \sum_j \sqrt{\frac{F_I^{\text{Music}}(j)}{\sum_k F_I^{\text{Music}}(k)} \cdot \frac{F_I^{\text{Motion}}(j)}{\sum_{k \in D} F_I^{\text{Motion}}(k)}} \quad (18)$$

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

5.1 実験データ

本研究では舞踊動作のデータベースとして、ブレイクダンス、インディアンダンスといったデータを用いて実験を行った。これらのデータはすべて CMU Graphics Lab Motion Capture Database から得たものである。データはすべて光学式モーションキャプチャシステムにて取得されたものであり、フレームレートは 120fps であった。また音楽データは CD から WAV 形式でキャプチャしたものである。実験に用いたデータはおよそ 60 秒で、サンプリングレートは 16bit, 44kHz であった。

5.2 音楽解析の結果

最初に音楽解析の結果を示す。リズムトラッキングを、クラシック、ロック、ジャズ、タンゴなど様々な種類の音楽データ 13 クリップに適用し、そのうち 10 の音楽クリップでリズムトラッキングが成功した。表 1 にその一部を示す。また、盛り上がり解析に関しても、盛り上がっていると感じるところで音楽インテンシティ特徴量が大きく現れ、うまく盛り上がりを追従できたといえる。また、音楽の構造解析に関してはジャズの曲のみうまくいかないものがあつた。これはジャズによくみられるアドリブ演奏によるもので、繰り返し構造が現れなかったことが原因と見られる。

5.3 舞踊動作生成の結果

図 7 に楽曲 “*T-Connection*” を用いて生成した舞踊動作の結果を示す。リズム成分、インテンシティ成分ともに楽曲と同期が取れているのがわかる。また生成された舞踊動作は、連結動作の解析を行っているため不自然な動きにならないものとなつ

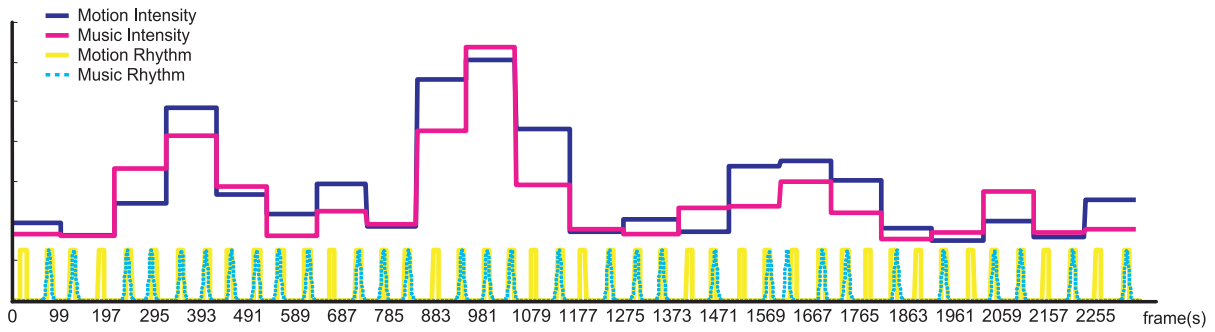


図9 “La Cumparsita” の音楽特徴量と生成された動きの特徴量の比較

黄：音楽リズム特徴量 緑：動きリズム特徴量 赤：音楽インテンシティ特徴量 青：動きインテンシティ特徴量

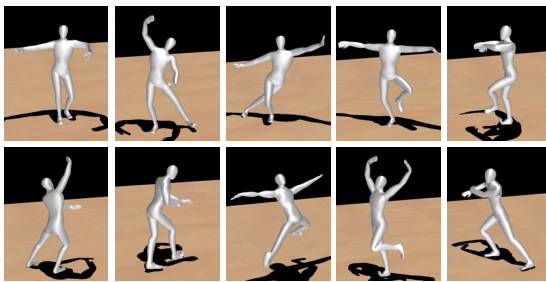


図7 “T-Connection” に合わせて生成された舞踊動作

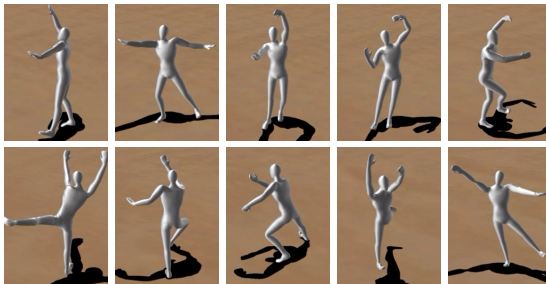


図8 “La Cumparsita” に合わせて生成された舞踊動作

ている。

図8に楽曲“La Cumparsita”を用いて生成した舞踊動作の結果を示す。また図9に、入力した楽曲を解析して得られた特徴量と生成された動きの特徴量の比較を示す。音楽特徴量と相関の取れた特徴量を持つ動きセグメントが抽出されていることがわかる。

5.4 計算コスト

本手法では、連結可能性の解析が連続するセグメント間すべてで評価値を求めるため、時間のかかる処理となっている。しかし Pentium-4 2.8GHz の PC を用いて 10 分ほどで処理できるため、実用可能なレベルの処理時間である。

6. おわりに

本稿では、入力した音楽データと相関の取れた舞踊動作を生成する手法を新たに提案した。具体的には音楽データおよび動きデータからリズム特徴量とインテンシティ特徴量を求め、それらがうまくマッチングするような動きセグメントを抽出、つなげ合わせることで新たな舞踊動作を自動で生成するというも

のである。結果からも、本手法の有効性が示された。今後の課題として、動きに現れる表情や表現といった特徴量を抽出する解析手法を開発し、音楽心理学と組み合わせてこれらを導入することを検討している。

謝 辞

本研究の一部は、日本学術振興会特別研究員制度および科学技術振興事業団戦略的基礎研究推進事業 (CREST) による支援を受けています。また、本研究で用いたモーションキャプチャデータは CMU Motion Capture Database (<http://mocap.cs.cmu.edu>) から使わせていただきました。ここに感謝の意を表します。

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] T. Shiratori, A. Nakazawa and K. Ikeuchi: “Detecting dance motion structure through music analysis”, Proc. IEEE Int'l Conf. on Automatic Face and Gesture Recognition, pp. 857–862 (2004).
- [6] R. Laban and L. Ullmann: “Mastery of Movement”, Princeton Book Company Publishers (1960).
- [7] T. Nakata, T. Mori and T. Sato: “Analysis of impression of robot bodily expression”, Journal of Robotics and Mechatronics, **14**, 1, pp. 27–36 (2002).
- [8] J. C. Brown: “Calculation of a constant Q spectral transform”, Journal of Acoustic Society of America, **89**, 1, pp. 425–434 (1990).
- [9] M. Wang, 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).
- [10] 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).
- [11] T. Kailath: “The divergence and Bhattacharyya distance measures in signal selection”, IEEE Trans. on Communication Technology, **COM-15**, pp. 52–60 (1967).