

加速度センサによるユーザコントロールを導入した音楽に合った舞踊動作の自動生成

安永卓哉^{†1} 中澤篤志^{†2,†3} 竹村治雄^{†2}

近年、キャラクターアニメーションに関する研究が数多く行われ、エンターテインメントの分野に応用されている。我々は従来からモーションキャプチャデータベースを利用して音楽に合った舞踊アニメーションを生成する手法を提案してきた。本稿ではこの手法を拡張し、ユーザの動きの盛り上がり加速度センサを用いて取得し動作生成に導入することで、インタラクティブにキャラクターの動きの盛り上がりコントロールできるシステムを開発した。音楽データからは楽曲構造解析によって曲をセグメントに分割し、リズム特徴量を抽出する。モーションキャプチャデータからは、音楽のリズムと高い相関を示す動きを候補セグメントとして取り出す。舞踊生成時は候補セグメントから、加速度センサによる盛り上がりの相関および隣接する候補セグメント同士の連結性をリアルタイムで考慮することで最適な動きを選択する。この手法を用いて、加速度センサによりインタラクティブに舞踊動作をコントロールできるアプリケーションを開発した。

Dance to Music Character Animation with accelerometer based user control

YASUNAGA TAKUYA,^{†1} NAKAZAWA ATSUSHI^{†2,†3}
and TAKEMURA HARUO^{†2}

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 from musical signals. We

first extract candidates of motion segment sets whose features are matched to those of music segments. For synthesizing dance performance, we select the motion segment set whose intensity by accelerometer and connectivity is matched to that of music segments. We evaluate all synthesizing time and matching value.

1. はじめに

リアルな動きを持つ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生成していると考えられる。音楽に合わせた動作を自動生成する手法はいくつか提案されており、個性的な動きである歌唱時の身振り¹⁰⁾や、音楽特徴量と動きの特徴量の学習による舞踊動作を生成する手法¹¹⁾がある。

しかし、従来の手法では、ユーザコントロールが導入されていないため、ユーザが舞踊動作の制御を行うことはできなかった。そこで今回は、ユーザの身体動作に着目し、加速度センサによってユーザの動きの盛り上がりを取得し、生成される舞踊動作の盛り上がりコントロールする手法を提案する。さらに、動作生成をリアルタイムで行うことで、インタラクティブな舞踊動作のコントロールシステムを開発した。本手法は以下の4つのステップから構成されている。

- (1) 動作解析
- (2) 音楽解析
- (3) 解析結果に基づく動きセグメント候補の抽出
- (4) 加速度情報に基づいた動作生成

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

†1 大阪大学 大学院情報科学研究科
Osaka University Graduate School of Information Science and Technology

†2 大阪大学サイバーメディアセンター
Cybermedia Center, Osaka University

†3 科学技術振興機構さきがけ
PRESTO, JST

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。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の成分(以後、インテンシティ成分)を抽出する。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。解析結果に基づき、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。加えて、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、動きセグメント間に遷移情報を付加する。動作生成においては、加速度情報に基づいて舞踊動作の盛り上がりをコントロールする。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り、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ることはしばしばある。このような楽曲のムードや盛り上がりを舞踊動作に反映させるため、加速度センサにより人の体の動き情報を取得し、人の動きの盛り上がりを舞踊動作の盛り上がりに反映させる。加速度の変化と動きセグメント候補との関連を求め、最適と評価された動きセグメントをリアルタイムに連結していくことで、舞踊動作を生成する。この舞踊動作生成に対して、オフライン処理に要する時間を計測し、加速度の変化と動きセグメントのインテンシティのマッチング評価を行った。

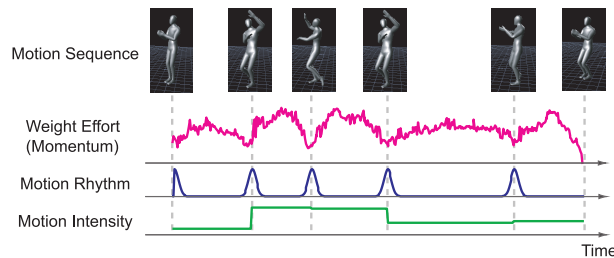


図 1 動きリズム特徴と動きインテンシティ(激しさ)特徴量

2. 動き特徴解析

我々の動作解析においては Laban の提案した Effort と呼ばれる動きの要素のうち、Weight Effort が非常に重要な役割を担っている。Weight Effort として人体の関節角度の線形和を用いる。

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

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。こうして得られた Weight Effort 成分 $W(f)$ から極小値を求め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関連をとるために用いる。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」成分との関連評価に用いる。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生成するとき重要な音楽特徴量として、我々は以下に示す 2 つの特徴量が特に重要である⁵⁾。

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。

音楽の構造解析および特徴量の算出のために、CQT⁸⁾で算出した各時刻におけるスペクトルパワーを用いている。構造解析では、CQT 特徴量ベクトルの各要素の自己相関を計算することで、繰り返しパターンを抽出する。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。リズム解析では、Onset Component(発音成分)からビート時刻を推定する。これは発音成分のピークと音楽のリズムが関連するからで

ある。

それぞれの特徴の求め方の詳細は論文⁵⁾を参照いただきたい。本研究では、音楽データから、音楽セグメントとリズム特徴量を取得し、利用する。

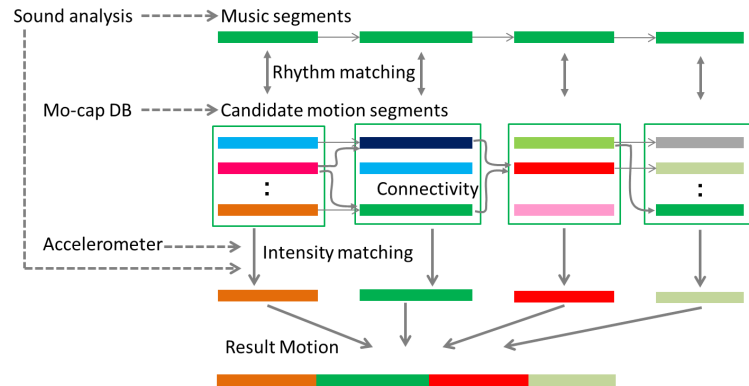


図2 舞踊動作生成アルゴリズム

4. 加速度によるユーザコントロールを導入した舞踊動作生成

得られた動き特徴量と音楽特徴量、および加速度センサによるユーザの動きを考慮して舞踊動作を生成する。ここで、加速度センサによるインタラクティブな舞踊動作生成を実現するためには、加速度センサにより得られた値によって、リアルタイムで次の音楽セグメントの動きセグメントを選択しなければならない。そのため、各音楽セグメントに合う動きセグメントを抽出しておく必要がある。

まずあらかじめ音楽データをオフラインで解析し、各音楽セグメントに合う動きセグメントと動きセグメント同士の遷移情報を保持した、モーションキャプチャデータの動作遷移データベースを作っておき、動作生成時にはリアルタイムで加速度センサの動きにあう動作を探索・連結することで、ユーザの動きに即座に対応した舞踊動作を生成する手法を開発した。図2にその概略を示す。オフライン処理では、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し、各音楽セグメントに対応する動きセグメント候補列を生成する。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

となる候補に連結可能性の遷移情報を付加する。動作生成時では、加速度センサより得られた加速度値の変化と動きセグメント候補との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リアルタイムで選択し出力とする。

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合う動きセグメントを抽出する。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。データベース内のすべての動きデータに対し、音楽セグメントのリズム特徴とモーションキャプチャから得られた動きリズム特徴との正規化相関をとり、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き、動きセグメント候補列を生成する。また、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。

4.2 動きセグメントの連結可能性評価

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、動作セグメント間でそれらが自然に連結するか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。得られた評価は各動きセグメントに遷移可能性情報として付加される。

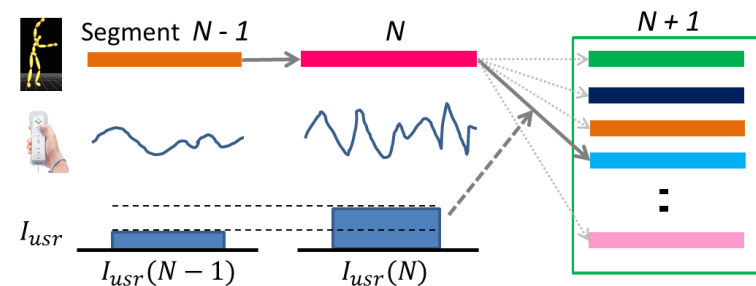


図3 リアルタイム動きセグメント選択アルゴリズム

4.3 加速度情報を用いたリアルタイム動きセグメント選択

動作生成時には、複数存在する動きセグメント候補の中から、加速度センサから得られたユーザの動きの激しさ（ユーザインテンシティ）の変化と相関の取れた動きセグメントをリアルタイムで選び出す。図3に概略を示す。\$N\$ 番目のセグメントの開始、終了フレーム数を \$t_s^N, t_e^N\$ とし、\$N+1\$ セグメントの選択は終了フレームの直前で開始する。これは、\$N+1\$ セグメント開始までに動きセグメント選択を終了する必要があるからである。また選択後、選ばれたセグメント列と現在のセグメント間の補間を行うため、音楽セグメントの間隔に応じてマージンをとって選択を開始している。

ここで、セグメント \$N\$ でのユーザインテンシティは以下のように定義する。

$$I_{usr}(N) = \sum_{t=t_s^N}^{t_e^N} (a_x^2 + a_y^2 + a_z^2) \quad (2)$$

ただし、\$a_x, a_y, a_z\$ は加速度センサから得られた \$x, y, z\$ 方向の加速度データである。

まず、現在の舞踊動作である \$N\$ 番目の動きセグメントに付加された遷移可能性情報から、次に連結可能である \$L\$ 個の動きセグメント候補のインテンシティ値 \$I_{mc}^1(N+1), I_{mc}^2(N+1), \dots, I_{mc}^L(N+1)\$ を抽出する。ここで、現在の動きセグメントと \$l\$ 番目の動きセグメント候補間のインテンシティ差は以下で得られる。

$$I_m(N) - I_{mc}^l(N+1) \quad (3)$$

これらから、動きセグメント間のインテンシティ変化とユーザインテンシティの変化が類似する \$l\$ を求める。

$$\underset{l}{\operatorname{argmin}} |(I_{mc}^l(N+1) - I_m(N)) - \alpha \cdot (I_{usr}(N) - I_{usr}(N-1))| \quad (4)$$

ただし、\$\alpha\$ は正の定数である。これにより、ユーザの動きの変化と動きセグメントのインテンシティ変化が最もよく相関する動きセグメントを選択できる。

4.4 ディスプレースメント・マッピングによる舞踊動作の生成

選択された動きセグメントに対してディスプレースメント・マッピング¹²⁾を行い、インテンシティを調整する。モーションキャプチャにおけるディスプレースメント・マッピングとは、全体の動きは保つが、個々の関節の動きを変化させることで、モーションキャプチャデータを編集する手法である。ユーザインテンシティが大きく下がった場合、動作が緩やかになるようななどのディスプレースメント・マッピングを行う。

図4にある関節に対するディスプレースメント・マッピングを示す。選択された動きセグメントの開始、終了フレームを \$f_s, f_e\$ とし、ある関節におけるモーションデータを \$m_{f_s}, \dots, m_{f_e}\$

とすると、モーションデータの始点から終点を結ぶ直線は以下で得られる。

$$f(i) = i \cdot (m_{f_e} - m_{f_s}) / (f_e - f_s) + m_{f_s} \quad (5)$$

ただし、\$i\$ はフレーム数である。この直線とモーションデータとの差を、縮小・拡大することでモーションデータを編集する。

$$f(i) + d \cdot (m_i - f(i)) \quad (6)$$

ただし、\$d\$ はユーザインテンシティによって変化する定数で以下に定義する。

$$d = \begin{cases} 1 & I_{usr}(N) - I_{usr}(N-1) < Threshold \\ 1 + \beta \cdot (I_{usr}(N) - I_{usr}(N-1)) & otherwise. \end{cases}$$

ただし、\$\beta\$ は正の定数である。\$d < 1\$ の場合、直線に沿ってモーションデータが縮小され、ゆったりとした動きになる。\$d > 1\$ の場合、直線に沿ってモーションデータが拡大され、大きな動きになる。

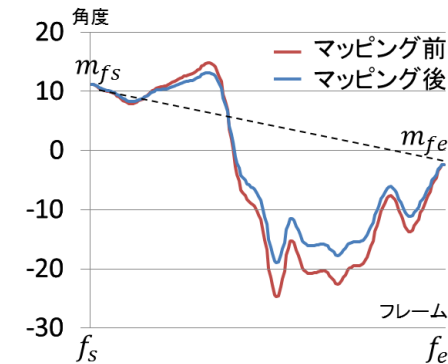


図4 ある関節に対するディスプレースメント・マッピング

4.5 連結動作の生成

最後に、選ばれた動きセグメント列と現在のセグメント間を補間し連結する。姿勢に関しては、各リンクの方向ベクトルをスプライン補間する手法を用いる⁵⁾。

5. アプリケーションの実装

本研究では、生成した舞踊動作をリアルタイムで表示するために、動作生成と同時に舞踊動作を表示するアプリケーションの実装を行った。実装したアプリケーションでは、舞踊動作の生成と音楽の時間同期を取り、生成のタイミングと音楽がずれないように実装した。図5はアプリケーションを実行した時の図である。

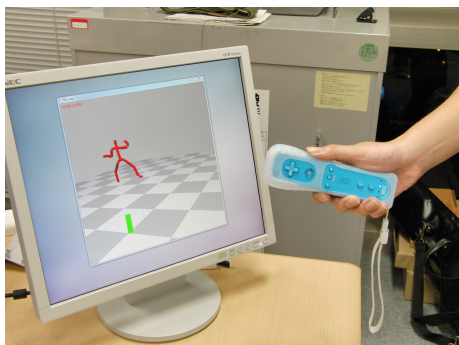


図5 アプリケーションの実行画面

本アプリケーションでは、加速度センサに Wii リモコン¹³⁾を使用した。Wii リモコンとは、任天堂株式会社から発売された家庭用ゲーム機 Wii¹⁴⁾の専用コントローラである。3次元加速度センサを内蔵しており、Bluetooth を用いて計算機と接続することで加速度値データを容易に取得できる。Wii リモコンを使用した理由は、安価であり容易に加速度値を取得でき、また身体動作を直接、舞踊動作の盛り上がりには反映させるためである。また、このアプリケーションでは、舞踊動作は BVH 形式で生成されている。BVH 形式とは、キャラクターの姿勢を人体モデルの各関節のオイラー角で表したもので、様々なモーションキャプチャデータベースや 3D グラフィックスソフトで対応している。

入力は音楽データとその音楽の動作遷移データベースとし、音楽の再生に同期し、各セグメントの動きセグメントの選択が順次行われていく。生成された BVH 形式の舞踊動作データはセグメントごとに表示プログラムに渡す。表示プログラムでも、音楽に同期をとって表示を行うことで音楽に合った舞踊動作がリアルタイムで表示される。

6. 実験

評価実験を行い手法の有効性を確かめた。モーションキャプチャデータ取得には、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。

6.1 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6)。それぞれ 3~5 曲のテンポや雰囲気の違いに合わせて、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。

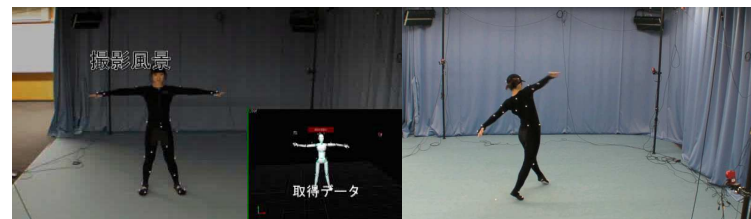


図6 実験データの取得シーン

曲名	長さ (sec.)	音楽解析	遷移情報データベース作成	合計
セントレイ	31	1.217	51.479	52.696
video killed the radio star	57	3.208	97.501	100.709
september	91	7.269	143.208	150.477

表1 オフライン処理に要する時間

6.2 動作生成に要する時間

オフライン処理に関する処理時間 表1は音楽ファイルに対して音楽解析、遷移情報データベース作成を行ったときにかかった処理時間を表したものである。57秒のwavファイルに関して、1つの音楽セグメントあたり平均1319個の動きセグメントが抽出され

ていた。

動作生成時の処理時間 動きセグメント候補列から加速度によりセグメントを選択し、連結動作を生成し、モーションを完成させる処理にかかる時間は平均して 0.0068(sec.) 程度である。十分に早い速度で処理を行えているため、リアルタイムで舞踊動作を生成することが可能である。しかし、動作補間生成のため音楽セグメントの終了前に 0.5 秒程度のマージンをとっているため、加速度による生成動作への反映が 1 セグメント分遅れる可能性がある。

6.3 加速度センサと動作の相関評価

加速度値の変化と選択された動きセグメントのインテンシティのマッチングを評価する。図 7 は青線が動きのインテンシティ値を、赤線が加速度の相対変化値を表している。加速度の変化と共に動きセグメントのインテンシティがうまく相関していることが確認できる。

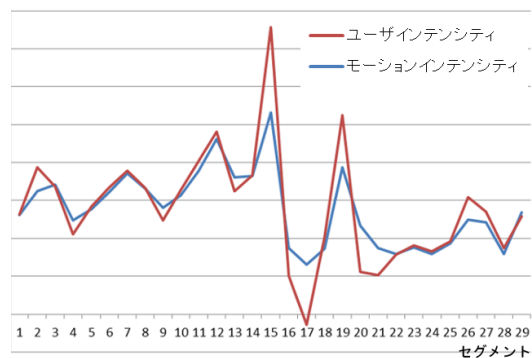


図 7 モーションのインテンシティと加速度値の相関評価

7. おわりに

本論文では、音楽からその曲に合った舞踊アニメーションを、モーションキャプチャデータベースを利用して生成する手法を述べ、ユーザが加速度センサを用いて動きの盛り上がりを実時間でコントロールすることを可能にした。また、生成に必要な時間とインテンシティのマッチング評価を行い、リアルタイムで舞踊動作を生成できることを確認した。今後の課題として、連結可能な動きセグメント候補がユーザインテンシティと合わない問題の解

消や、動きセグメント音楽特徴量・動き特徴量の学習、ユーザインテンシティを改良し、より直観的なコントロールを目指すことが挙げられる。

謝辞 本研究は、JST 戦略的創造研究推進事業さきがけの支援を受けておこなわれました。

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. Popovi'c: " 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) "The Dance to Music Charater Animation Project", <http://d2m.ime.cmc.osaka-u.ac.jp/>, 2009.
- 10) 青塚寛之, 山本正信, 歌唱時の身振りと音楽との関連性に基づく音楽からの動作生成, 電子情報通信学会論文誌, Vol.J90-D, No.11,pp.3055-3064, 2007
- 11) Rukun Fan, Songhua Xu and Weidong Geng, Example-based Automatic Music-Driven Conventional Dance Motion Synthesis. IEEE Transactions on Visualization and Computer Graphics , 2011
- 12) A.Bruderlin and L.Williams, Motion signal processing. Computer Graphics (Proceedings of SIGGRAPH 95), 29,97–104, 1995
- 13) NINTENDO,Wii Remote, 2006
- 14) NINTENDO,Wii, <http://www.wii.com/>, 2006.