

# ダンスコラボレーションのための 隠れマルコフモデルによるダンス動作の識別

尾崎良太<sup>1</sup>, 鶴田清也<sup>1</sup>, 崔雄<sup>2</sup>, 八村広三郎<sup>3</sup>

<sup>1</sup> 立命館大学 理工学研究科 情報理工学専攻, <sup>2</sup> 立命館大学 グローバル・イノベーション研究機構,

<sup>3</sup> 立命館大学 情報理工学部

本研究は、実ダンサーとバーチャルダンサーとがダンスコラボレーションを行うために、リアルタイムに実ダンサーの動作を識別することを目的としている。ここでは、モーションキャプチャによるダンサーの動作データから特徴量を抽出し、その特徴量に対してベクトル量子化を行い、隠れマルコフモデル(HMM)によりダンス動作の識別を行う。7種類のブレイクダンス動作を用いて実験を行い、ダンスコラボレーションに使用可能な程度の識別率が得られるどうかを検討した。この結果、90%程度の識別率を得た。また、ダンスコラボレーションにおいてリアルタイムに識別を行うことに対応するために、いくつかの動作が連続する一連の動作が入力された場合の対応方法についても考察した。

## Recognition of Dance Motions Using Hidden Markov Model for Dance Collaboration

Ryota Ozaki<sup>1</sup>, Seiya Tsuruta<sup>1</sup>, Woong Choi<sup>2</sup>, Kozaburo Hachimura<sup>3</sup>

<sup>1</sup> Graduate school of Science and Engineering, Ritsumeikan University. <sup>2</sup> Global Innovation Research Organization, Ritsumeikan University. <sup>3</sup> College of Information Science and Engineering, Ritsumeikan University.

### Abstract:

The purpose of this research is to recognize dance motions in real time for dance collaboration between a human dancer and a virtual dancer. We extracted feature values from motion data of the human dancer using a motion capture system. We classified dance motions with vector quantization to the feature values and Hidden Markov Models. We considered the possibilities of using for dance collaboration by experiments of recognition intended for 7 break dance motions. We considered the approach to the recognition of successive several motions for solving the problem which will happen in a real time application.

### 1 はじめに

映画やテレビゲームなどのエンタテインメント分野、動作解析やリハビリなどの医療分野において、モーションキャプチャシステムやVR技術が広く用いられるようになってきている。本研究室ではこのモーションキャプチャシステムとVR技術を活用し、人と人、もしくは人とバーチャルダンサーとのインタラクティブなダンスコラボレーションを行うシステムについての研究を行っている[1]。

図1にコラボレーションシステムの構成を示す。このシステムにおいて、リアルタイムに実ダンサーの動作を識別し、それに応じてバーチャルダンサーが動作を返すことで、バーチャルダンサーとのインタラクティブ

なやり取りを体感することができる。実ダンサーの動きに応じて、動作が変化するバーチャルダンサーとのコラボレーションによって、実ダンサーにとっても観客にとっても、興味深いコラボレーションシステムを実現する。

[1]の研究では、動作データから、頭、両手、両足の速度の平均と標準偏差と腰のy座標の速度の平均と標準偏差などの特徴量を抽出し、その特徴量から主成分分析を用いて、学習データを作成し、主成分得点を算出してダンス動作の識別をリアルタイムに行う。しかし、この手法は特徴量の時系列変化に基づくものではなく、70%程度の識別率しか得られていない。この識別率ではスムーズなコラボレーションが行えるとは言い難く、さらに識別率を上げ、80%から90%程度の識

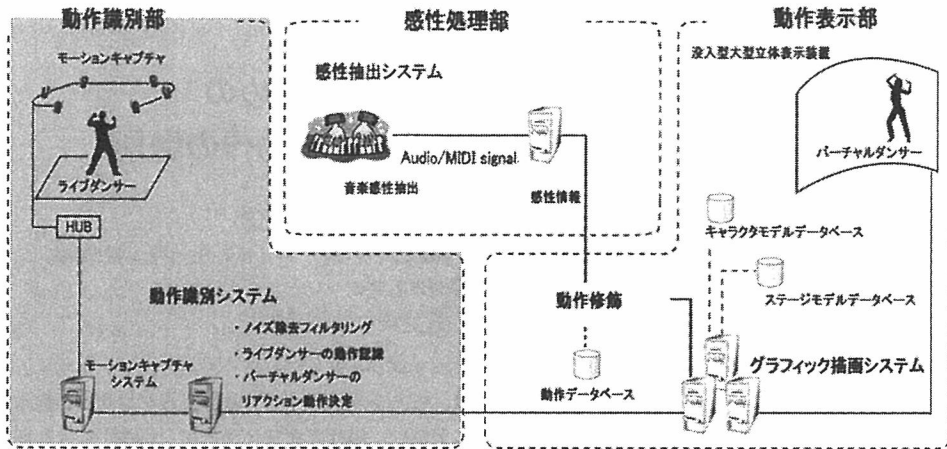


図 1: ダンスコラボレーションシステム

別率が得られる手法が必要である。

[1]では、学習データ主成分分析による主成分得点に基づく識別手法をとっているので、不特定者の（学習データに用いていない人も含めた）ダンス動作の識別には不向きである。また、上述のようにダンス動作の時系列変化を考慮した識別はできない。

そこで本研究では、時系列変化を考慮したモデルである、隠れマルコフモデル (Hidden Markov Model:HMM) を用いたダンス動作識別手法を提案する。

HMM は確率モデルを用いたマッチング法で、学習可能な手法であり、時間軸方向の伸縮に強いという長所があるため、音声認識、動画像認識、ジェスチャ認識などの時系列パターン認識において広く用いられている [2][3]。他にも時系列パターンの認識では、テンプレートマッチング法である DP マッチングも広く用いられているが、最尤推定によって認識ができ、複数の学習データからの学習ができる HMM の方が、本研究のような不特定者を対象にした場合に優位であることが確認されている [4]。

ここでは、本手法がダンスコラボレーションに使用できるかどうかを検討するために、7種類のブレイクダンスの動作を用いた識別実験を行い、実際にコラボレーションを行う際に生じる問題についても考察した。

## 2 HMMによる動作認識

### 2.1 HMMの概要

HMM はマルコフ過程に基づく確率的な時系列パターンの生成源のモデルである。その中には相互に遷移する複数の状態とその生起確率および、状態間の遷移確率が与えられており、出力確率に従って、各状態から遷移する際に、シンボルが出力される。このシンボルの系列がこのモデルの生成するパターンとして考えられ、これらのパターンの属するクラスごとに1つのHMMモデルが構成され、これらを用いて未知パターンの認識が行われる。以下、図2に従ってHMMの動作について述べる。

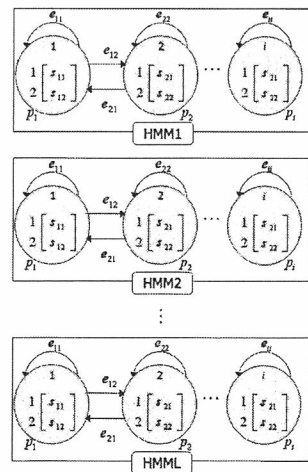


図 2: HMM の例

ここでは、1と2の出力シンボルが観測可能とする。楕円1, 2, ...,  $i$ は状態とその系列を表しており、 $s_{ij}$ は状態 $i$ の時にシンボル $j$ を出力する確率、 $p_i$ は状態 $i$ の初期確率、 $e_{mn}$ は状態 $m$ から状態 $n$ へ遷移する際の状態遷移確率を表している。Lクラスのデータを認識する際には、このようなHMMをL個作る。

以上のHMMのパラメータは既知パターンを用いた学習を行うことで決定する。

クラスが既知の入力パターンが出力する複数のシンボル列に対して、後述するアルゴリズムを利用してこれらのシンボル列を出力するように各種の確率パラメータを調整し、このクラスのための1つのHMMを構成する。このHMM学習過程を、対象とするそれぞれのクラスのパターンについて行い、クラス数分のHMMを構築する。

次に、学習を行ったHMMを用いて、未知パターンによる最尤推定を行う。

例えば、クラスが未知の入力パターンが出力するシンボル列が「121」の場合、すべてのHMMにおいて、シンボル列「121」の出力確率を求め、比較することで尤度を最大とするHMMが求められる。各HMMはそれぞれのクラスに対応しているので、HMMを特定することにより、未知パターンがどのクラスに所属するかを認識できる[5]。

HMMはある状態からすべての状態に遷移できる全遷移型モデルと、状態遷移が一定方向に進むleft-to-rightモデルに大きく分けられる。あらかじめセグメンテーションされていることが多く、時間的に非可逆である音声認識では、left-to-rightモデルが多く用いられるが、本研究ではセグメンテーションされていない身体動作データを使用するため、そのデータの認識に適している全遷移型モデルを使用する[6]。

さらにHMMは各状態における出力確率分布の表現の違いにより、離散型分布モデルと連続分布型モデルと半連続型モデルに分けられる。本研究では、リアルタイム性を重視するため、計算量の少ない離散型分布モデルを使用する[6]。

## 2.2 ベクトル量子化

離散型分布モデルでは、入力された時系列データを有限個の離散シンボルで構成されるシンボル列に変換する必要がある。そのための前処理としてベクトル量子化手法を用いる。ベクトル量子化は、

1. M個のN次元ベクトルデータの集合をK個のクラスにクラスタリングする。

2. 各クラスの代表ベクトルを求めて、そのクラス内のデータを代表ベクトルで置換する。

3. 代表ベクトルを符号化(シンボル化)する。

という手順で行われる。

クラスタリングには、一般的にLBGアルゴリズムが用いられ、代表ベクトルと符号の対応関係を記したコードブックが作成される。LBGアルゴリズムは、繰り返し2分割とk-meansアルゴリズムを組み合わせることで実現され、以下の様な手順で行われる。

1. ベクトル集合を2クラスに分割
2. 各クラス内の重心ベクトルを求める
3. 各クラスに属するベクトルと重心ベクトルの差が閾値より小さくなるまで1.と2.を繰り返す
4. クラス数がK個になれば、クラスタリングを終了する。それ以外なら1.に戻る。K個になった時の各クラスの重心ベクトルを代表ベクトルとする。この場合のKは、2のべき乗となる。

例えば図3に示すように、2次元空間内にベクトル1~14が存在し、これを4種類にベクトル量子化手法によって符号化する。

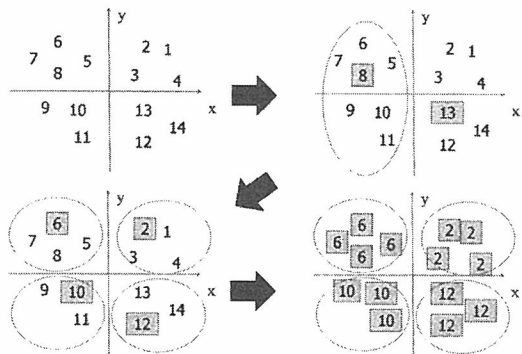


図3: ベクトル量子化

## 2.3 HMMにおける学習

HMMにおける学習では、一般的に以下のようにBaum-Welchアルゴリズムが用いられる。

1. 出力確率と状態遷移確率に適当な初期値を割り振った初期HMMを作成する。

2. 初期 HMM において、学習シンボル列のあらゆる可能な状態遷移系列に対して、生成確率 (シンボル出力確率×状態遷移確率) を計算する。
3. それぞれの状態遷移系列に対して、シンボル出力確率と状態遷移確率を求め、それらに対して 2 で求めた生成確率を足し合わせたもので、パラメータを更新する。
4. この操作をパラメータの更新量が閾値より小さくなるまで繰り返す。

## 2.4 HMM における認識

HMM における認識では、認識対象のシンボル列を出力する確率 (尤度) が最大となるモデルを求める。この際、本研究のようにリアルタイムでの認識を目的とする場合、すべての状態系列の確率を計算するのを避けるため、一般的に以下のような Viterbi アルゴリズムが用いられる。

Viterbi アルゴリズムは、ある時点の遷移で、確率の高くなる方だけを計算し、それ以外の系列は計算を打ち切ることによって、最も確率が高くなる状態遷移系列だけを求めて近似する手法である。このアルゴリズムを用いて求めた系列におけるシンボル出力確率と状態遷移確率から、尤度を計算し、一番尤度の高い HMM に対応するものを認識結果とする。

## 3 ダンス動作識別システム

本研究におけるダンス動作識別システムの概要を図 4 に示す。システムは、学習部と識別部から構成される。

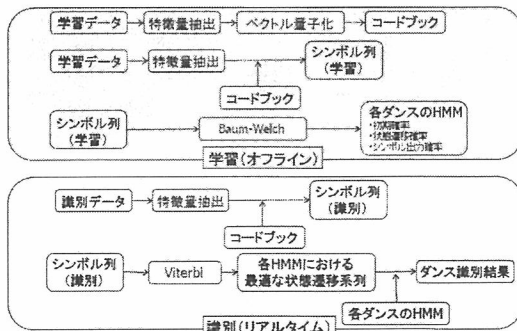


図 4: ダンス動作識別システムの概要

## 3.1 学習

学習部では、あらかじめモーションキャプチャシステムを用いて取得しておいたユーザのダンスモーションデータを学習データとする。その学習データから特徴量を抽出し、これを基に、ベクトル量子化を行い、コードブックを作成する。作成したコードブックを用いて、最近傍法によって学習データをシンボル化し、シンボル列を作成する。このシンボル列から Baum-Welch アルゴリズムを用いて、ダンス動作の種類ごとの HMM を作成する。

### 3.1.1 特徴量抽出

モーションキャプチャで取得したデータから、各ダンス動作の特徴量を抽出する。ここでは、図 5 に示すようなスライディングウィンドウによる手法を用いる。現在のフレームより 1 つ前のフレームから時間をさかのぼる方向に幅 W フレーム分のウィンドウを設定し、このウィンドウ内の動作データを現在フレームの特徴量を抽出するために用いる。

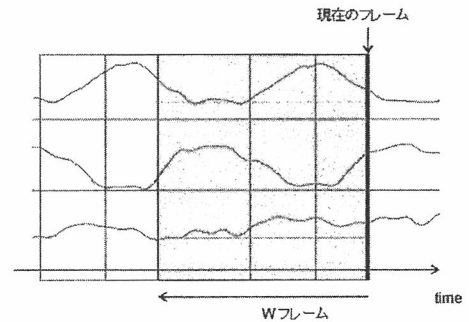


図 5: スライディングウィンドウ

ここでは、このウィンドウ内での頭 (HDTOP)、両手 (RWRA・LWRA)、両足 (RHEE・LHEE)、腰 (ROOT) の各マーカの速さの平均値と標準偏差、腰のマーカの y 座標の平均値と標準偏差の、合計 14 個の特徴量を用いる。用いたマーカの位置を図 6 に示す。

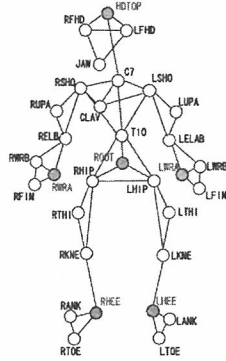


図 6: マーカの位置

### 3.1.2 ベクトル量子化とコードブック作成

本研究では、各ダンス動作の 1 サイクルに相当するウィンドウ幅での特徴量を各 4 つずつ用意し、それらを学習データとしてベクトル量子化を行い、コードブックを作成する。ここで、クラス数を 16 とした時のコードブックの例を図 7 に示す。

符号(クラス)	代表ベクトルの要素			
1	602.015	260.275	...	915.7533
2	607.64	215.7	...	939.495
3	865	413.74	...	850.49
4	957.32	512.415	...	853.385
5	1223.775	480.04	...	904.58
6	1095.38	513.4	...	918.31
7	819.97	274.69	...	969.775
8	1272.12	459.31	...	910.68
9	729.15	422.7	...	482.9
10	811.49	474.97	...	470.82
11	1088.37	766.87	...	493.9
12	1071.87	734.77	...	473.86
13	1013.5	540.61	...	289.9
14	1135.905	559.84	...	299.605
15	1024.54	553.28	...	325.51
16	855.4682	406.1361	...	763.3225

図 7: 作成したコードブックの例

### 3.1.3 シンボル列 (学習) の作成

作成したコードブックを用いて、学習データからシンボル列を作成する。

1. あるスライディングウィンドウ幅で、学習データから特徴量を抽出し、抽出した特徴量とコードブックに記された代表ベクトルの要素に対して最近傍法を用いる。
2. 特徴量を一番距離に近い代表ベクトルの要素に対応した符号に置き換え、その符号をシンボルとみなす。

3. あるフレーム数分のシンボルを得るまで、1, 2 を行い、得られたシンボルからシンボル列を作成する。

以上を各学習データに対して行い、シンボル列 (学習) を作成する。

### 3.1.4 各ダンス動作の HMM の作成

作成したシンボル列から Baum-Welch アルゴリズムを用いて、各ダンス動作ごとに HMM を作成する。図 8 にダンス動作 Entry の HMM を例として示す。この HMM は 16 種類のシンボルと状態数 3 のパラメータを持ち、ある 1 つの Entry の学習データから得たシンボル列を用いて学習を行ったものである。

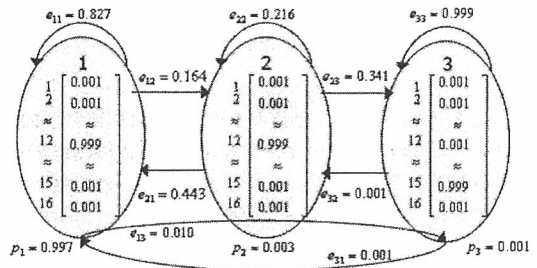


図 8: 作成した HMM の例

## 3.2 識別

識別の際には識別部で、モーショキャプチャシステムから送られてくるダンス動作のストリーミングデータを対象として扱う。このストリーミングデータを識別データとし、学習と同じ手法で特徴量の抽出とシンボル列 (識別) の作成を行う。

シンボル列 (識別) と学習で作成した各 HMM に対して、Viterbi アルゴリズムを用いて、シンボル列 (識別) を出力する最適状態遷移系列を求める。

この系列におけるシンボル出力確率と状態遷移確率から、シンボル列 (識別) を出す確率を計算し、最大の確率を与える HMM を求め、これに対応するダンス動作を識別結果として出力する。

## 4 動作識別実験と実験結果

2Step, 6Foot, Bronx, Entry, Shift, TopRock, UpRock の 7 種類のブレイクダンスの基本ダンス動作を対象として用いて、実験を行った。各ダンス動作のスナップショットを図 9 に示す。

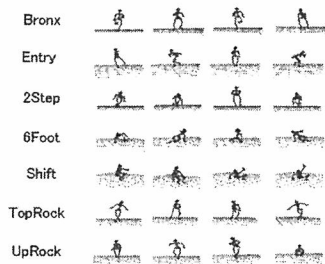


図 9: 用いたダンス動作

ここでは、各ダンス動作の 1 サイクルに相当する 120 フレーム分のスライディングウィンドウ内の動作データから特徴量を抽出し、シンボル化する。この際のフレームレートは 60 フレーム/秒とする。シンボル化の際、シンボル数 ( $K_S$ ) を 16 と 32 の 2 つの場合を用意し、16 クラスと 32 クラスの場合のコードブックを作成した。学習データとして各動作 4 個ずつのデータを用意した。このデータから学習シンボル列の長さ ( $L_S$ ) が 60 と 120 のシンボル列を作成し、状態数 3 の HMM をダンス動作ごとに作成した。

次に、同じダンサーのものであるが、学習で用いたものとは別のデータを各動作 10 個ずつ用意し、識別シンボル列の長さ ( $R_S$ ) を 60 と 120 に設定した。

以上、 $K_S$ ,  $L_S$ ,  $R_S$  各 2 つずつの組合せ、合計 8 パターンのパラメータを用い、各動作それぞれを識別データとして入力した場合に、その動作を識別できるかどうかについての識別実験を行った。パラメータのパターンごとの識別結果を図 10 に示す。

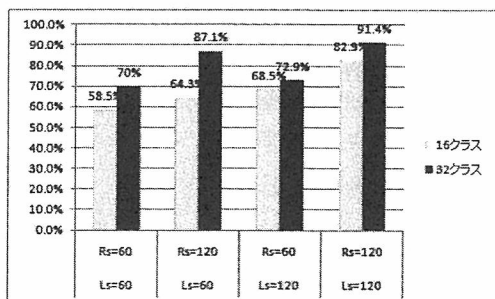


図 10: パターンごとの比較

この結果、 $K_S = 16$ ,  $L_S = 60$ ,  $R_S = 60$  の場合が最も低い識別率を与え、 $K_S = 32$ ,  $L_S = 120$ ,  $R_S = 120$  の場合が最も高い識別率になった。また、各々のパターンを比較しても、 $K_S = 16$  よりも  $K_S = 32$ ,  $L_S = 60$  よりも  $L_S = 120$ ,  $R_S = 60$  よりも  $R_S = 120$  の方が高い識別率を得られた。

この結果より、16 のシンボルではクラスタリングが適切に行えないことがわかる。

ダンス動作の識別において良い識別率を得るには、ある程度の長さの学習シンボル長と識別シンボル長が必要であることも分かる。 $L_S = 60$  は時間で表すと 1 秒に相当する長さであり、これは対象としたダンス動作を識別するには短すぎる時間であったと考えられる。

以上のことから、対象としたブレイクダンスの基本ダンス動作を識別する際は、 $K_S = 32$ ,  $L_S = 120$ ,  $R_S = 120$  というパラメータを用いるのが良いと考えられる。もちろん、このパラメータは対象とするダンス動作に依存するので、他の動作の場合には別途検討する必要がある。

最悪の識別率と最良の識別率を得られたパラメータでの各ダンス動作の識別結果を表 1、および表 2 に示す。

表 1: 最悪の識別結果 ( $K_S=16, L_S=60, R_S=60$ )

		出力							Total
		Bronx	Entry	2Step	6Foot	Shift	TopRock	UpRock	
入力	Bronx	30	20	0	0	0	0	0	
	Entry	50	50	0	0	0	0	0	
	2Step	20	10	70	0	0	0	0	
	6Foot	0	0	0	100	0	0	0	
	Shift	0	40	20	0	40	0	0	
	TopRock	0	40	10	0	0	50	0	
	UpRock	0	80	0	0	0	0	20	
	Total								58.5

表 2: 最良の識別結果 ( $K_S=32, L_S=120, R_S=120$ )

		出力							Total
		Bronx	Entry	2Step	6Foot	Shift	TopRock	UpRock	
入力	Bronx	70	10	10	0	10	0	0	
	Entry	0	100	0	0	0	0	0	
	2Step	0	0	100	0	0	0	0	
	6Foot	0	0	0	100	0	0	0	
	Shift	0	0	0	0	30	10	0	
	TopRock	0	0	10	0	10	80	0	
	UpRock	0	0	0	0	0	0	100	
	Total								91.4

表1に示した  $K_S = 16$ ,  $L_S = 60$ ,  $R_S = 60$  の場合、すべてのダンス動作において、これらは Entry に誤識別されることが多い。これは、 $K_S = 16$  の場合、クラスタリングが適切に行えなかったため、Entry の HMM で出力確率が高い「15」というシンボルがすべての学習データで多く出現したことが原因であった。

表2に示した  $K_S = 32$ ,  $L_S = 120$ ,  $R_S = 120$  の場合、抽出した特徴量が他の動作と似ていた識別データにおいて誤識別が起こっていたため、特徴量の選択や特徴量の抽出する際の時間区間に改善の余地があると考える。

Bronx に関してだけ、 $K_S = 16$ ,  $L_S = 60$ ,  $R_S = 60$  の場合の方が高い識別率になっているが、Entry や 2Step が Bronx と誤識別されることも起こるため、全体の識別率を考えると、対象としたブレイクダンスの基本ダンス動作を識別する際のパラメータとしては適していない。

## 5 連続識別による識別

### 5.1 連続識別手法による識別実験

4章の実験では、図11に示すように、 $R_S$  が 60、または 120 に達した時点で識別結果を出力し、識別を終了した。そのため、それ以降に正しい識別結果と判断される部分が存在しても、その部分は識別が行われないため、誤識別が起こることがあった。

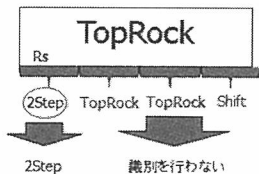


図 11: 誤識別例

そこで、図12のように連続して同じ動作と識別された場合に、その動作を識別結果として出力するという連続識別手法を用いることで、さらに識別率を上げられるのではないかと考え、この方法を試みた。

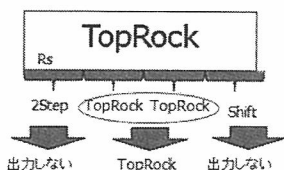


図 12: 提案手法

今回は、4章の実験と同条件下で、2回連続して同じ動作であると識別された場合を、識別結果として出力することにして、識別実験を行った。

$K_S = 32$  の場合に対して、この手法を適用する前(4章の実験結果)と後の識別率の比較を図13に示し、最良の識別率を得られたパラメータでの各ダンス動作の識別結果を表3に示す。この時のパラメータは、 $K_S = 32$ ,  $L_S = 120$ ,  $R_S = 60$  と  $K_S = 32$ ,  $L_S = 120$ ,  $R_S = 120$  であり、各ダンス動作の識別結果も同じであった。

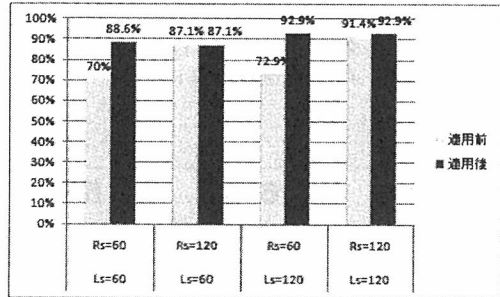


図 13: 手法適用前と後の比較

表 3: 手法適用後の識別結果 ( $K_S = 32$ ,  $L_S = 120$ ,  $R_S = 60$  と  $K_S = 32$ ,  $L_S = 120$ ,  $R_S = 120$ )

		出力							Total
		Bronx	Entry	2Step	6Foot	Shift	TopRock	UpRock	
入力	Bronx	70	10	10	0	10	0	0	
	Entry	0	100	0	0	0	0	0	
	2Step	0	0	100	0	0	0	0	
	6Foot	0	0	0	100	0	0	0	
	Shift	0	0	0	0	90	10	0	
	TopRock	0	0	0	0	10	90	0	
	UpRock	0	0	0	0	0	0	100	
	Total								92.9

この結果より、本手法を適用することで、識別率を向上できることが分かった。また、今回は実験的に2回連続という設定を用いたが、それ以上の回数やある区間内での多数決による決定などが適している可能性も考えられるため、今後の実験で検討する必要がある。

### 5.2 一連動作の識別への利用

ここまで各ダンス動作を対象とした場合の識別について述べたが、実際のダンスコラボレーションでは、1

つのダンス動作だけがくり返し行われるのではなく、数種類の異なるダンス動作が一連の動作として行われ、その一連動作がストリーミングデータとして送られてくる。本研究では、ダンスコラボレーションを行う際に、リアルタイムにダンス動作を識別することを目的としているため、この一連動作の識別に対応する必要がある。

一連動作を識別する際、ダンス動作が切り替わるところでは2つの動作が混合した特徴量が抽出されるため、正しい識別結果を得ることができない可能性が高い。これには5.1で述べた、連続識別手法を利用することで対処できるのではないかと考えている。

未実装ではあるが、この手法を利用した場合の連続動作の識別のイメージ図を、図14に示す。

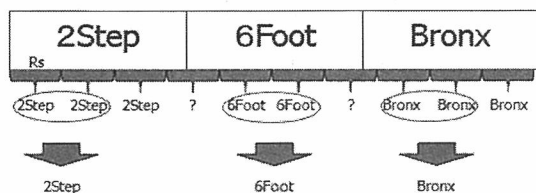


図 14: 一連動作の識別のイメージ

## 6 まとめと今後の課題

本研究では、ダンスコラボレーションシステムを構築するために、HMMによるダンス動作の識別を試みた。これにより、ブレイクダンスの基本ダンス動作を対象としたダンス動作識別実験を行った。

この実験から、各ダンス動作をそれぞれ1つのストリーミングデータとして送った場合は、90%程度の識別率を得られることが分かった。さらに、連続識別による識別手法が識別率向上に繋がることも分かった。

今後は、この手法を利用して一連動作の識別実験を行い、その結果によっては、特徴量の選択、特徴量を抽出する際の時間区間、HMMの状態数などを検討する必要があると考える。この一連動作の識別において、80%から90%程度の識別率を実現することで、ダンスコラボレーションシステムに十分使用できるシステムの実現を目指す。

## 謝辞

本研究の一部は、文部科学省科学研究費基盤研究(B)19300031、および、文部科学省グローバルCOEプログラム「日本文化デジタル・ヒューマニティーズ拠点」の援助により行われた。ここに記して謝意を表する。

## 参考文献

- [1] 鶴田清也, 川内大和, 崔雄, 八村広三郎, パーチャルダンスコラボレーションシステムのための実時間動作認識, 映像情報メディア学会誌, Vol62, No6, pp.909-913, 2008.
- [2] 大和淳司, 上田修功, 和田俊和: 動作認識のための状態遷移モデル, 人工知能学会誌, 17(1), pp.41~46, 2001.
- [3] 豊倉行崇, 南角吉彦, 後藤富朗, 徳田恵一, 北村正: 基本動作 HMM に基づく手話単語認識手法, 電子情報通信学会基礎・境界ソサエティ大会講演論文集, pp.72, 2006.
- [4] 春山智, 小林哲則: 動画像処理による手振り動作認識, 電子情報通信学会総合大会講演論文集, No.2, p362, 1997.
- [5] 荒木雅弘: フリーソフトでつくる音声認識システム, 森北出版株式会社, 2007.
- [6] 村上仁一, 杉山 雅英, 渡辺 秀行: Ergodic HMM を用いた未知・複数信号源クラスタリング問題の検討, 電子情報通信学会論文誌, Vol.J78-D-2 No.2, pp.1997-2004, 1995.