

概要

日本の伝統舞踊や匠の技といった人の動きの文化財，すなわち無形文化財は人から人へ継承することで保存，伝承されてきた。しかし近年では，後継者の減少などに伴い散逸が起こりつつあり，それらをデジタル化して保存する技術が求められている。

そこで，本研究はモーションキャプチャシステムを用いて人の動きをデジタル化し，舞踊動作を直接解析することにより有限の動作プリミティブを抽出，そして動作プリミティブを用いて舞踊動作を記述することを目標としている。動作プリミティブとは動きを構成する最小単位のことを示す。本論文では，これまでの手足の速度や関節角の角速度のみを用いた手法に加え，音楽を考慮した動作解析手法を提案する。これは，動作プリミティブ間にはとめ動作と呼ばれるキーポーズが存在すること，またほとんどの踊りが音楽のリズムに合わせて踊られる，という知識に基づくものである。実際には，音響信号からビートトラッキングを行い，そのリズムに基づいてセグメンテーションを行うことにより，動作シーケンスからキーポーズを取り出して，舞踊の構造を理解する。

目次

第 1 章	序論	1
1.1	本研究の背景と目的	2
1.2	本論文の構成	6
第 2 章	関連研究	7
2.1	ビートトラッキングに関する関連研究	8
2.1.1	ビートトラッキングとは	8
2.1.2	ビートトラッキングの研究例	9
2.2	モーションキャプチャシステムとその応用	11
2.2.1	モーションキャプチャシステムとは	11
2.2.2	一般動作を対象とした解析・合成に関する研究	13
2.2.3	舞踊動作を対象とした解析・合成に関する研究	14
2.3	本研究の位置づけ	15
第 3 章	舞踊音楽のビートトラッキング	18
3.1	リズムを構成する音楽的要素	19
3.2	発音時刻の検出	20
3.3	リズムの開始時刻と平均間隔の算出	22
3.3.1	平均リズム間隔の算出	22
3.3.2	リズム開始時刻の算出	25
3.4	ビートトラッキング	26
第 4 章	動作プリミティブの抽出	28
4.1	動作プリミティブとは	29
4.2	解析の対象となる動き	30
4.2.1	解析の対象となる身体の部位	30
4.2.2	解析に用いる座標系	31
4.3	セグメンテーション候補点の算出	33

4.3.1	手・重心の場合	33
4.3.2	足の場合	33
4.4	身体の各部位のセグメンテーション	35
4.5	動作プリミティブの抽出	36
第5章	実験	38
5.1	実験データ	39
5.2	ビートトラッキングの結果	39
5.2.1	リズム間隔推定の誤差評価	43
5.2.2	舞踊音楽のリズム推定結果	44
5.3	動作プリミティブ抽出の結果	47
5.3.1	女性の踊り手による会津磐梯山踊りの場合	49
5.3.2	男性の踊り手による会津磐梯山踊りの場合	49
5.3.3	じょんがら節の場合	50
第6章	結論	59
	謝辞	62

目次

1.1	「標準日本舞踊譜」で提案されている舞踊譜の例	4
1.2	Labanotation の例	5
2.1	光学式モーションキャプチャシステム	11
2.2	磁気式モーションキャプチャシステム	12
2.3	アルゴリズムの概要図	17
3.1	発音時刻の検出	21
3.2	スペクトログラムと各周波数帯発音時刻の例	23
3.3	自己相関関数	24
3.4	リズムのひずみに対応したビートトラッキング	27
4.1	重心座標のモデル	31
4.2	体中心座標系	32
4.3	足や重心のスピード	32
4.4	手・重心のセグメンテーション候補点算出の流れ	34
4.5	手・重心のセグメンテーション候補点の算出方法	34
4.6	足のセグメンテーション候補点の算出方法	35
4.7	身体の各部位のセグメンテーション	36
5.1	Vicon Motion Systems で使われるマーカー	40
5.2	会津磐梯山踊り	41
5.3	じょんがら節	42
5.4	ビートトラッキング結果 - 会津磐梯山踊り	45
5.5	ビートトラッキング結果 - じょんがら節	46
5.6	女性による会津磐梯山踊りのキーポーズ抽出結果	52
5.7	女性による会津磐梯山踊りの結果の比較	53
5.8	男性による会津磐梯山踊りのキーポーズ抽出結果	54
5.9	男性による会津磐梯山踊りの結果の比較	55

5.10	じょんがら節のキーポーズ抽出結果(一部)	56
5.11	じょんがら節のキーポーズ抽出結果(全て)	57
5.12	じょんがら節の結果の比較(一部)	58

表 目 次

5.1 wav ファイルのパラメータ	39
5.2 平均リズム間隔の推定値とその誤差	43
5.3 踊り手にとっての会津磐梯山踊りのキーポーズ	47
5.4 踊り手にとってのじょんがら節のキーポーズ	48
5.5 女性の踊り手による会津磐梯山踊りの結果比較	49
5.6 男性の踊り手による会津磐梯山踊りの結果比較	50
5.7 じょんがら節の結果比較	51

第1章

序論

1.1 本研究の背景と目的

本研究は大きく分けて 2 つの背景から成り立っている。

散逸しつつある伝統舞踊の保存

日本の伝統舞踊や匠の技といった人の動きの文化財、すなわち無形文化財は人から人へ継承することで今まで保存、伝承されてきた。歌舞伎や能、狂言などでは伝承体制が確立されている一方で、地域に残る民族舞踊では後継者が減少し、その文化が今まさに消えようとしている。こういった流れの中、ユネスコでは世界遺産として各国を代表する伝統芸能を登録し、保護の対象としており、また日本国内においても日本国憲法に日本古来の文化を守るという一文を加えるかどうかの議論が始まりつつある。こういった動向から、無形文化財を保存し、いつでも時空間的に再現できるよう符号化して保存する方法を確立することが急務となってきている。

音楽には時間軸方向で各楽器のピッチがどう変化するか、またその音をどのくらいの時間発し続けるか、などについて示した楽譜が存在する。この楽譜のように、舞踊譜と呼ばれる身体の動きを記述する方法について、伝統舞踊の保存を目的として様々な提案がなされている。東京国立文化財研究所(現独立行政法人東京文化財研究所)では「標準日本舞踊譜」[1]を刊行し、伝統舞踊を記述・保存するための舞踊譜の規格を統一しようと試みた。この本に掲載されている舞踊譜の例を図 1.1 に示す。しかし、3次元空間の莫大な情報量を紙面に記述する作業は非常に煩雑であり、またその舞踊譜では舞踊のすべてを表現することができなかつたため、結果として普及しなかつた。

一方、バレエなど西洋の舞踊教育現場では Labanotation と呼ばれる舞踊譜 [2] がよく用いられている。Labanotation での舞踊の記述例を図 1.2 に示す。西洋舞踊と日本の伝統舞踊での相違点を克服し、Labanotation を日本の伝統舞踊に適用できるように改良している研究もある [3]。しかし Labanotation はあくまで舞踊教育用であるため、手先の軌道などの表現に関する細かい描写に関してはほとんど記述がすることができず、舞踊の保存や再現といった目的にはあまり適していないと言える。

こういった背景から我々は、舞踊という 3次元空間での表現を記録するためには、モーションキャプチャシステムを利用して動きをデジタル化し、CG やロボットなどを用いて擬似的にもしくは実際に再現できる方が、人間の直感に訴えやすく理解し

やすいのではないかと考えている。ビデオなどの映像に収めればよいのではないかという意見もあるが、映像では任意視点からの映像を得ることができず、隠れなどによって見にくい場面が多々ある。そこで我々は、モーションキャプチャを使って伝統舞踊を保存することを目指している。モーションキャプチャシステムは身体の関節部などにマーカーをつけ、赤外線反射や磁場の歪みなどからマーカーの位置を読み取って記録することで、人の動きを再現できるシステムである。しかし、舞踊動作を記録する際、取得した動きデータでは被験者の動きを忠実に再現するだけであり、舞踊動作のキーポーズや動作素などを知ることができない。伝統舞踊を保存するためにはキーポーズや動作素を抽出して舞踊動作をシンボリックな記述に変換し、そして舞踊をCGなどで再現する際には、このシンボリックな記述から舞踊動作へ変換、生成できることが望ましい。本研究はこのような条件を満たす舞踊動作のシンボリックな記述方法の提案を目的とした、キーポーズの抽出、および動作シーケンスのセグメンテーションの方法について提案する。

「踊る」知能をロボットへ

本研究では舞踊動作をシンボリックな記述に変換するだけでなく、ヒューマノイドロボットによって舞踊を再現することも目的としている。これは、CGよりヒューマノイドロボットを用いて舞踊を模倣する方が理解しやすいということに由来するものである。

舞踊と音楽は非常に密接な関係がある。人間は音楽を聴いてリズムを感じ取り、そのリズムに合わせて身体を動かすことで舞踊を表現しているからである。つまりヒューマノイドロボットが音楽を無視して身体を動かすだけでは、人間の舞踊を完全に模倣しているとは言いがたいと言える。多くの舞踊動作では音楽のリズムに合わせてとめ動作と呼ばれるキーポーズが存在する。このキーポーズを理解しておくことで、ロボットが音楽と同期を取って身体を動かすことができれば、舞踊を模倣したといえると考えている。この機能をサウンドフィードバックと呼ぶことにし、サウンドフィードバックを実現して舞踊を模倣するロボットを最終目的として、本研究ではその最も根本的な問題であるキーポーズと音楽の関係の推定方法を提案する。

廓 八 景 坂 東 流

【本調子】

	チン 4	トン 0	チャ 4 0
右手	① 扇		
左手	扇ヲ取り 本 構		
右足	立 チ 東		
左足			

窄メ扇ヲ前ニ置き

板付 常座り 礼

ラン 4	チン 6 ツン 4	チン 0 ナン 3	チ 4 ン
ス	■		—
み	わ	た	せ ば
②			
	>		
	扇 取		
	>		
	引掛り		

チン 6	チン 4	チ 6 ン	リン 6 ス
■		—	●
			は
③			
	>		
	本 構		
	>		
	東寄セ		

チ 6	リ 6 ハ	リ 6 ハ	リ 6 ハ
—			—
	な	の	
④			
	>		
	抜出シ		

図 1.1: 「標準日本舞踊譜」で提案されている舞踊譜の例

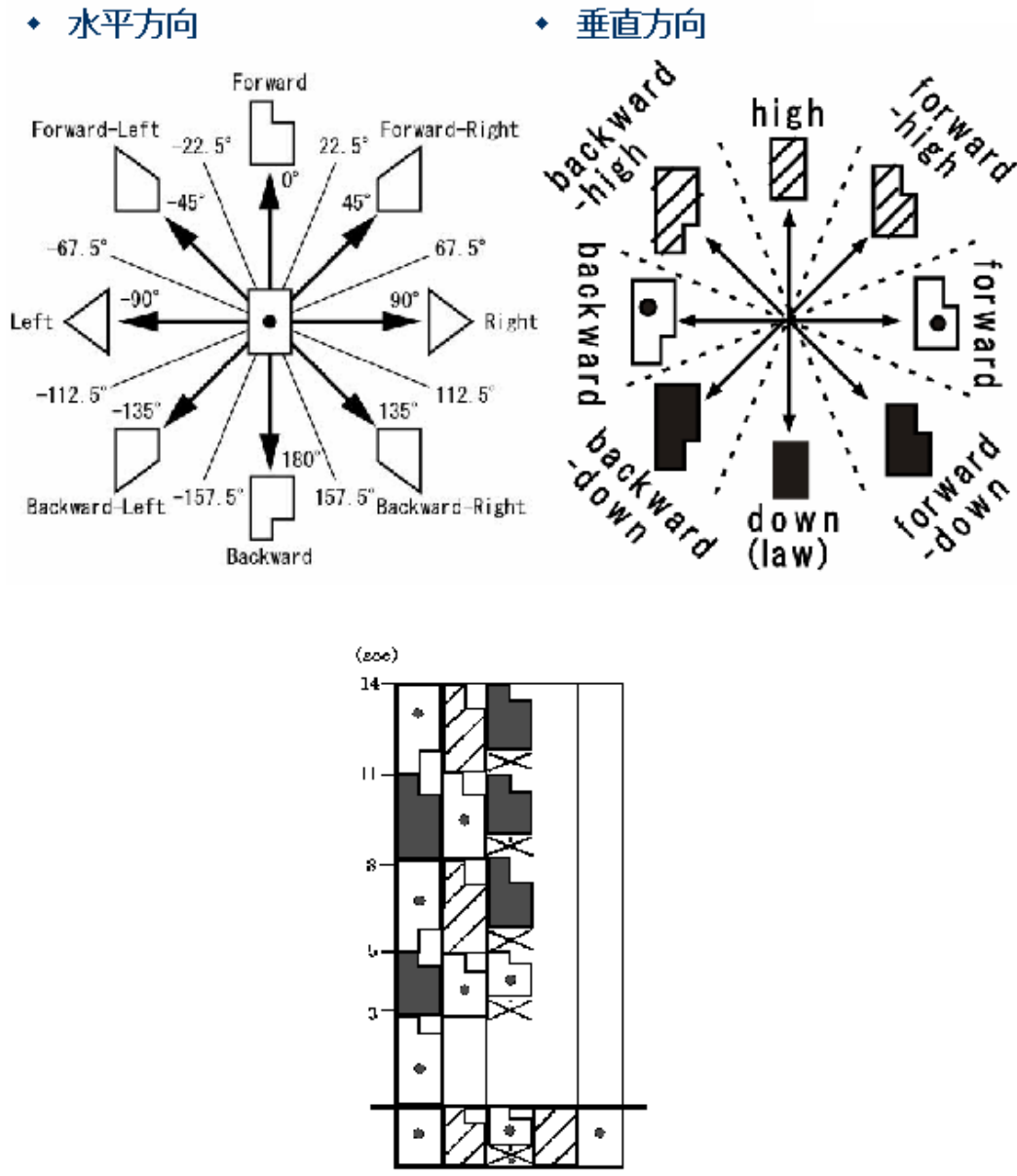


図 1.2: Labanotation の例
 上段：身体の各部位が示す方向の記述の仕方
 下段：Labanotation における舞踊譜

1.2 本論文の構成

第 1 章は本章であり，本研究の背景と目的，概要について述べた．第 2 章では，モーションキャプチャとその応用に関する研究例，および音楽情報処理において重要なビートトラッキングの研究例について述べる．第 3 章では，本研究で提案したビートトラッキングの手法について述べる．第 4 章では，ビートトラッキングによって推定された音楽のリズムに基づいて，舞踊動作を解析してキーポーズを抽出する手法について提案する．第 5 章ではこれらの手法の実験とその結果を示し，評価を行なう．第 6 章では，本論文で述べた研究結果をまとめるとともに，その結果について検討を行なう．また今後の課題についても述べる．

第2章

関連研究

本章ではビートトラッキングに関する関連研究，およびモーションキャプチャシステムとその応用例について述べ，そこから本研究の位置づけについて考察する。

2.1 ビートトラッキングに関する関連研究

2.1.1 ビートトラッキングとは

人間は流れてくる音楽に合わせて手拍子を打ったり，歌ったり，演奏したりすることができる。これはすなわち音楽に対する時間方向の構造化の認識を行なう知能が人間にあることを示している。この知能をコンピュータで実現しようとする研究がビートトラッキングであり，音楽聴取過程をコンピュータ上で実現するための第 1 段階として重要であるとみなされている。

西洋音楽の時間軸方向の構造は，拍(ビート)→小節→楽句→楽節→楽曲というように階層的なものになっている。東洋音楽と西洋音楽とでは楽理が異なるため，東洋音楽では小節などの概念があるとは限らないが，拍単位でのリズムが存在することは確かである。音楽が演奏・生成されるときについて考えてみると，演奏者の頭の中で思い描くビート構造に従って時間軸方向に調和した演奏がなされる。このことから，このビート構造を演奏音中に示す行為を音楽を生成する際の順問題(入力から結果として示される出力を予測する問題)ととらえることができ，また演奏された音楽(出力)から演奏者の頭の中にあるビート構造を推定するビートトラッキングは逆問題(出力結果から入力を推定する問題)と捉えることができる。

順問題に相当するビート提示行為では，演奏者はビート構造を特定の音でビート構造を明示的に表現しているわけではなく，様々な音楽的要素の関係の中に非明示的に表している。しかもその音楽的要素は一定ではなく，ジャンルや楽曲などによって異なることが多い。例えば，打楽器音の含まれたポピュラー音楽ではバスドラムやスネアドラムのパターンの中にビート構造が示される傾向があり，スイングジャズの曲ではライドシンバルのパターンにビート構造が示される傾向が見られる。

ビートトラッキングが本質的に難しいのは，音楽中に非明示的にしか表現されていない階層的なビート構造を音楽から推定する逆問題であるからである。したがってビートトラッキングの難易度は，どれくらい明示的にビート構造が表現されているかで決まり，単純に楽器数などでは決まらない。ただし入力対象が，各演奏音の

発音時刻, 音高, 音量, 音長が容易に得られる MIDI 信号ではなく, マイクを経由した空気振動のサンプリングによって得られる音響信号である場合, ビート提示行為, 楽器発音, 音響伝達からなる全処理過程の逆問題となるため, 音響信号からビートの手がかりとなる音楽的要素を抽出する問題もビートトラッキングに含まれる. 実際には楽器数が多く信号が複雑になっているケースが多いため, ヒューリスティックな知識を導入することで, その楽曲のジャンルなどに応じたビートトラッキング手法を提案している例も多い.

2.1.2 ビートトラッキングの研究例

ビートトラッキングの研究は入力信号が MIDI 信号か音響信号かによって大きく分かれる. これは MIDI 信号が各演奏音の発音時刻, ピッチ, 音量, 音長が容易に得られるのに対し, 音響信号は振幅値が記録されているだけで音楽的要素が直接記録されていないためである.

またリズムには単一レベルのビート構造と階層的なビート構造が存在し, そのどちらを推定するかによって方法が変わってくる. 一般的に単一レベルのビート構造が拍単位のリズムを示すのに対し, 階層的なビート構造は 4 分音符/2 分音符/小節といった構造のことを示す. 以下では入力信号の形式, 推定結果の種類ごとに過去の研究を紹介していく.

MIDI 信号を対象としたビートトラッキング

前述の通り, MIDI 信号はその仕様からより多くの音楽的要素を取得しやすく, 手がかりとしやすい. 基本的には, それらの音楽的要素からビート構造を推定するためのヒューリスティックな知識・ルールを導入し, 特徴的な音楽的要素の存在する時刻やそれが変化する時刻が, ビート時刻と一致するように推定することが多い.

まず単一レベルのビート構造を扱った研究として, A. Driesse の手法 [4] では各演奏音の発音時刻, ピッチ, 音量を用いたビートトラッキングが検討された. 曲が始まる時点でのテンポを与え, そのあとにあらかじめ設定しておいた優先度に応じてヒューリスティックルールを適用し, 次のリズム時刻を推定する.

R. Rowe は人間の演奏にインタラクティブに反応できるセッションシステムを構築し, その一要素としてビートトラッキングを用いた [5]. このシステムではマルチ

エージェントモデルが適用されており、各演奏音の発音時刻を扱うエージェントを中心に、コード情報を扱うエージェントなどとの相互作用を考慮して、ビートトラッキングを行なっている。

D. Rosenthal は和音演奏されたテンポ変化のあるクラシックのピアノ曲を対象としたビートトラッキング手法を提案している [6]。リズムの手がかりとしては発音時刻、ピッチ、音長、同時発音をまとめたコード、伴奏から分離されたメロディといった様々な音楽的要素と優先度付けされたヒューリスティックルールを活用して、小節レベルまでの階層的なビート構造を推定していた。

音響信号を対象としたビートトラッキング

音響信号では容易に音楽的要素を取り出すことができないので、主に発音時刻やそれに関連したパワー変化などを手がかりとすることが多い。そのため、MIDI 信号の場合と比べてヒューリスティックルールを導入する段階まで踏み込んだ研究例は少ない。

ピアノやギターといった単一楽器による和音演奏の音響信号を対象とした研究として、B. Vercoe は各周波数帯域ごとの発音時刻を算出し、自己相関関数にかけることで次のビート時刻が期待される度合いの分布を求める方法を提案した [7]。しかしこの分布から次のビート時刻を決定していく方法については議論されていなかった。

片寄は自動採譜システムの一部として階層的なビート構造を推定するビートトラッキングを実現した [8]。この手法では、入力のパワー変化の微分値を手がかりとして、パワー変化の微分値が大きいところがビート時刻になるように重み付けしてビートのパルス列を求めていた。また小節の頭ではパワーが大きいと仮定して小節単位の推定も行なっていた。

N. Todd は発音時刻を手がかりとしてリズムの繰り返し構造を推定する方法を提案している [9]。この繰り返し構造の推定には rhythmogram と呼ばれる表現を用いており、場合によっては小節レベルの繰り返し構造も推定できるとしている。

後藤はテンポがほぼ一定であるポピュラー音楽を対象として発音時刻、コードの変化、打楽器音の発音パターンを手がかりとして、小節レベルまでの階層的なビート構造を推定する手法について提案している [10]。このシステムでは打楽器音の有無の判別結果に応じて選択的に音楽的知識を切り替えることで、打楽器音がある場合とない場合をともに扱うことを可能にしている。

2.2 モーションキャプチャシステムとその応用

2.2.1 モーションキャプチャシステムとは

近年、映画やテレビゲームといった身近なメディアでコンピュータグラフィクスが使われるようになり、キャラクターをよりリアルにアニメーションさせることが非常に重要になってきている。人間らしい動きをソフトウェア上で作成するには多大な手間やコスト、時間がかかり効率的ではないため、解決策としてモーションキャプチャシステムが広く使われている。モーションキャプチャシステムとは、主に身体の関節部にセンサーを取り付けた被験者に動作を行ってもらい、そのセンサーの動きを追従してデジタルデータに変換し、PC 上で擬似的に動かすことができるシステムである。現在のモーションキャプチャシステムは主に光学式と磁気式の 2 種類の計測方式が広く普及している。



図 2.1: 光学式モーションキャプチャシステム

光学式のモーションキャプチャシステムでは、複数台のカメラから赤外線などの

不可視光を発射し、センサーの反射光を読み取ることで空間上でのセンサーの位置を把握するシステムになっている (図 2.1). 特徴としては、環境からの影響を受けにくいという長所を持つが、その構造上からセンサーの向きが把握できない、またマーカーが身体の一部に隠れてしまうオクルージョンが発生しやすいといった問題点がある.



図 2.2: 磁気式モーションキャプチャシステム

一方磁気式モーションキャプチャシステムでは、エミッターと呼ばれる装置で磁場を作り、その磁場の中でのセンサーの動きを読み取るシステムになっている (図 2.2). 磁場を利用するため、鉄筋建造物や屋外での使用ができない、キャプチャエリアを広く取れないといった問題点があるが、センサーが向いている向きも把握できるという長所がある.

このように、モーションキャプチャシステムは 3D アニメーションソフトなどと比べて、人間の動きを忠実に再現することができるため、非常に需要が伸びてきている. しかし実際にはアニメーターなどが望んでいる動きと、実際にモーションキャプチャシステムから得られる動きデータとは異なることが多く、動きデータを修正,

加工，合成する必要が出てくる．この修正や加工，合成を現在市販されているアプリケーションで行うには多大な手間や時間，スキルが必要になってしまう．こういった背景から動きデータの加工，修正して，より有効的に利用することを目的として，動きデータの解析・合成に関する研究が盛んに行なわれてきている．

2.2.2 一般動作を対象とした解析・合成に関する研究

動きデータを有効活用するためには，動きを解析して動きの特徴抽出やセグメンテーションなどを行い，その解析結果をもとに動きを合成することで，新しい動きを得ることを目的としたものが多い．

Armin Bruderlin らは信号処理で用いられている手法を応用することでこの目的を達成している [11]．具体的には，ローパスフィルタやバンドパスフィルタを用いた動きデータの周波数分解を行い，任意の周波数成分に重み付けをすることで，一つのモーションから複数のモーションを作成する方法や，DP マッチングを用いて二つの動きの同期を取り，合成する手法などを提案している．大崎らや中澤らは，手や足の速度の極小点，すなわち手足の動きが止まっているところを見つけ出し，その点で動きをセグメンテーションすることで動きの基本単位を見つけ出す手法を提案している [12][13]．求めた動きの基本単位を任意に組み合わせることによって新しい動きを生成している．また M. Brand ら [14] は，例えば歩行動作にも「速い歩き」「遅い歩き」など複数種類あるように，こういった一つのクラスに分類される動きでの様々な種類を隠れマルコフモデルで表す手法を提案している．Andrew Witkin らは，キャラクタの周りの環境情報を与えることで，元の動きデータから障害物に干渉しないような動きを作成する手法について提案している [15]．また障害物に限らず，たとえばテニスのラケットを振る動きを一つ用意をしておき，ボールの来る位置を指定することによって，そのボールの位置にあったラケットのスイング動作を生成することができる．今こういった手法は Motion Warping，または Motion Displacement Mapping と呼ばれている．

また，動きはキャラクタの体格や動作の所要時間によって左右されるものもある．例えば物を持ち上げる動作の場合，腕が短いキャラクタと長いキャラクタとでは，腕の伸ばし方，すなわち肩や肘の関節角が大きく異なってくる．このように，データ取得時の被験者の体格と CG キャラクタとの間に体格の差があるときは，得られたデータを CG キャラクタに適用できるよう最適化しなければならない．こういった

空間的な制約や時間的な制約をまとめて Spacetime Constraints と呼ぶ。

Michael Gleicher [16] や Jehee Lee [17] らは足の接地などの Spacetime constraints を考慮しつつ、キャラクタの体格にあった動きに加工する手法について提案している。Charles Rose ら [18] は、Spacetime Constraints をトルク最小化の問題として捉え、これを BFGS 法によって解くことで、計算コストを減らす手法について提案している。

以上の手法は、一つの動きを加工することで様々な条件に対応した新たな動きを生成していた。もう一方では、たくさんの動きデータをデータベースとしてもち、その中から条件に合った動きを合成手法についても多く提案されている。Lucas Kover らの手法 [19] では、データベース内のそれぞれの動作シーケンス間を遷移動作で結び、Motion Graph と呼ばれるグラフ構造をあらかじめ構築しておく。そしてユーザの入力条件に合うように Motion Graph 上を遷移していくことで、有限長の動作シーケンス群から任意長の動作シーケンスを作成することができる。Jehee Lee ら [20] は二層構造の Motion Graph を作成し、そのグラフ上の遷移によってユーザの入力した軌道に対応した動きを合成する手法について提案している。二層構造の Motion Graph のうち、下位層では各フレーム間での遷移確率を持ったマルコフモデルを適用し、上位層では姿勢の類似度に基づいたクラスタリングを行う。そしてユーザが指定した軌道において、キャラクタがとるべき姿勢を上位層のクラスタから導き出して大まかな動きを求め、下位層の確率に基づいてフレーム単位で姿勢を決めるというものである。Okan Arikan ら [21] は腰の位置や各関節に関する評価関数を設け、この値に基づいて Motion Graph を作成している。さらにこのグラフに、遺伝的アルゴリズムで使われる mutation を導入してノードを増やすことにより、より多くの種類の動きを合成することを可能にしている。

2.2.3 舞踊動作を対象とした解析・合成に関する研究

舞踊動作は以下の点で一般動作と大きく異なる。

- 音楽などの外部情報と同期を取っている
- 舞踊の解釈の違いや個人的な趣向から、個人差が如実に現れる

前者の特徴に着目した舞踊動作の解析手法として Tae-hoon らの研究 [22] が例として挙げられる。具体的には、舞踊や行進などの周期的な動作シーケンスにおいて、各

関節角の角速度の極小点を求め、極小点の現れる頻度から Motional Beat と呼ばれる動きのリズムを算出し、そのリズムに基づいてセグメンテーション、合成を行なうことで、リズムカルな動作を作ることができる。

後者の特徴に着目した舞踊動作の解析手法としては、中澤らの研究 [23] が挙げられる。この手法では動きが、個人差などに関係のない共通の動き = “Base Motion” と個人差などによる部分 = “Style” が組み合わさっているものと仮定し、それぞれを抽出する方法について述べられている。

舞踊教育においては、身体の動かし方のパターンを人から人へわかりやすく提示・教授したいというニーズが昔からあり、楽譜のようにシンボリックな記述で舞踊動作を表そうとする研究がいくつかある。バレエなどの西洋舞踊を対象とした舞踊の提示方法としては、Ann による Labanotation [2] が最も有名であり、現在でも多くの舞踊教育現場で使われている。八村ら [3] は西洋舞踊対象とした Labanotation を日本舞踊にも適用できるように拡張する研究を行なった。

一方中岡ら [24] は、舞踊は 3 次元空間上での表現であり Labanotation やコンピュータグラフィックスでの表現では十分な理解が得られないという考えから、ヒューマノイドロボットを用いて舞踊動作を提示する手法を提案している。

2.3 本研究の位置づけ

本研究の目的は舞踊動作の構造を理解するために、動作プリミティブと呼ばれる基本動作群を抽出することである。Flash ら [25] によると、動作プリミティブとは「人はすべての動作系列を記憶しているのではなく、基本単位から構成される」という仮定から来るものであり、この基本単位を動作プリミティブと呼ぶ。動作プリミティブを抽出できれば舞踊動作をシンボリックに記述することが容易になり、音楽における楽譜のようなものを舞踊動作で作成することが可能になると思われる。

また Flash らは動作プリミティブと動作プリミティブの間は手足が静止した状態になっているとも述べている。一般的に舞踊動作では「とめ動作」と呼ばれる全身が止まっている状態があり、踊り手もこのとめ動作をキーポーズとして認識していることから、このとめ動作を検出できれば動作プリミティブを抽出できるといえよう。前述の中澤らの手法 [13] や大崎らの手法 [12] では動きを手足の速さに基づいてセグメンテーションを行いシンボリックな記述で表す手法を提案しているが、得られた各セグメントの時間長がまちまちでリズムに即したのではなく、必ずしも舞

踊動作に適した手法とは限らない。また Tae-hoon Kim らの手法 [22] は、関節角の角速度の極小点間の間隔を元に動きのリズムを求め、それを元にセグメンテーションを行っており、リズムの要素を失ってはいない。しかしすべてのセグメント長を同一と仮定しているため、本来一まとまりとするべき動作を分割してしまうという問題点がある。本研究ではこういった問題点を解決した上でキーポーズの抽出、およびセグメンテーションによる動作プリミティブの抽出を行なうことを目的とする。本研究における提案手法は以下にあげる二つのステップに分かれる。

- 舞踊音楽のビートトラッキング
- 音楽のリズムを考慮した舞踊のキーポーズの抽出

舞踊音楽のビートトラッキングでは、まず音響信号を時間領域から周波数領域に変換し、演奏音が発せられたときのスペクトルパワーの立ち上がりを求める。そのスペクトルパワーからリズム開始時刻、平均リズム間隔を求め、最後にリズムのひずみ分を補正する。

一方舞踊のキーポーズの抽出においては、エンドエフェクターと重心座標をそれぞれ適した座標系に変換し、速さを求める。その速さから各部位のキーポーズの候補点を求め、推定された音楽のリズムを用いてキーポーズの候補点を絞り込む。そして、最後に簡単な論理演算を用いてキーポーズを推定、セグメンテーションにより動作プリミティブの抽出するという流れになっている。

このアルゴリズムの流れを図 2.3 に示す。

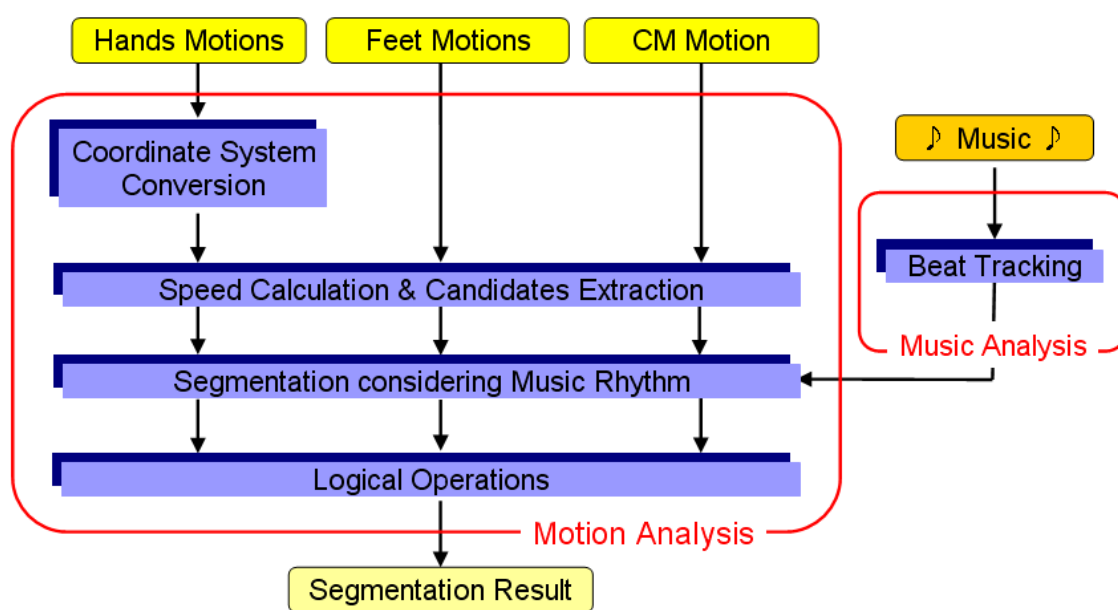


図 2.3: アルゴリズムの概要図

第3章

舞踊音楽のビートトラッキング

本章では、本研究で提案するビートトラッキングのアルゴリズムについて説明をする。

3.1 リズムを構成する音楽的要素

本手法では後藤の提案手法 [10][26] を元に、日本の舞踊音楽にもビートトラッキングを行なえるよう改良を加えた。

後藤の提案するビートトラッキングシステムでは以下の条件を想定している。

条件 1 ロックやジャズといった西洋音楽の楽典を元に作曲されたポピュラー音楽

条件 2 リズムは最初から最後までほぼ一定

本研究で対象とする日本の舞踊音楽においてもリズムは最初から最後までほぼ一定であり、条件が一致するためこの手法を参考にした。

また後藤は、こういった条件を満たす音楽におけるビートトラッキングの手がかりとなる音楽的要素として以下の3つをあげている。

発音時刻 音が鳴り始めるタイミング

コードの変化 長調・単調などの和音の構造の変化のタイミング

ドラムの発音パターン スネアドラムやバスドラムの発音パターン

これらの要素にヒューリスティックな知識を考慮することで高い精度でのビートトラッキングに成功している。

本研究では対象が西洋音楽ではなく日本の舞踊音楽であるため、まずはじめにこれらの音楽的要素が日本の舞踊音楽のビートトラッキングに適用できるかどうかを考察することから始めたい。

- 発音時刻
 - － 音が発せられた瞬間がリズムのタイミングと一致する可能性が高いということに関しては、音楽の種類を問わずに言えることであるため、有効

な手段の一つであると判断できる。また後藤の手法では発音時刻の検出方法として、スペクトルパワーが増加しているか否かで判定していることから、発音時刻は音楽の種類に依存せず検出することができる。

- コードの変化
 - 後藤の手法ではコードの変化の検出に周波数成分を用いている。しかし、西洋音楽における和音(コード)の周波数特性と問うよう音楽における和音の周波数特性は大きく異なるため、そのまま適用することは難しいと考えられる。本手法ではこの要素は利用しなかった。
- ドラムの発音パターン
 - 後藤の手法ではドラムの発音パターンの検出にはスネアドラムとバスドラムの周波数特性のテンプレートマッチングによって検出している。しかし日本の舞踊音楽で使われる打楽器の周波数特性とスネアドラムやバスドラムの周波数特性は大きく異なるため、本手法ではこの要素を利用しなかった。

以上から、本研究での日本の舞踊音楽のビートトラッキングには「発音時刻」のみを用いた。

また、後藤の手法では小節単位までの階層的なビート構造を推定することが目的であるが、東洋音楽と西洋音楽とでは根本的に楽典が異なるため、本研究の対象である日本の舞踊音楽では西洋音楽のような階層的なビート構造になっているとは限らない。そこで、本研究では単一レベルのビート構造、すなわち拍単位のリズムを推定することを目的とする。

3.2 発音時刻の検出

ある音が発せられたとき、その音に対応した周波数のスペクトルパワーが増加する。そのスペクトルパワーの増分を求めることによって発音時刻の検出を行う。ある時刻 t 、周波数 f において、スペクトルパワーの増加分を算出する関数 $d(t, f)$ を式 3.1 定義する。

$$d(t, f) = \begin{cases} \max(p(t, f), p(t+1, f)) - \text{PrevPow} & (\min(p(t, f), p(t+1, f)) \geq \text{PrevPow}) \\ 0 & (\text{otherwise}) \end{cases} \quad (3.1)$$

$$\text{PrevPow} = \max(p(t-1, f), p(t-1, f \pm 1)) \quad (3.2)$$

ここで $p(t, f)$ は時刻 t , 周波数 f におけるスペクトルパワーを表す. 図 3.1 に $d(t, f)$ 算出における概要図を示す.

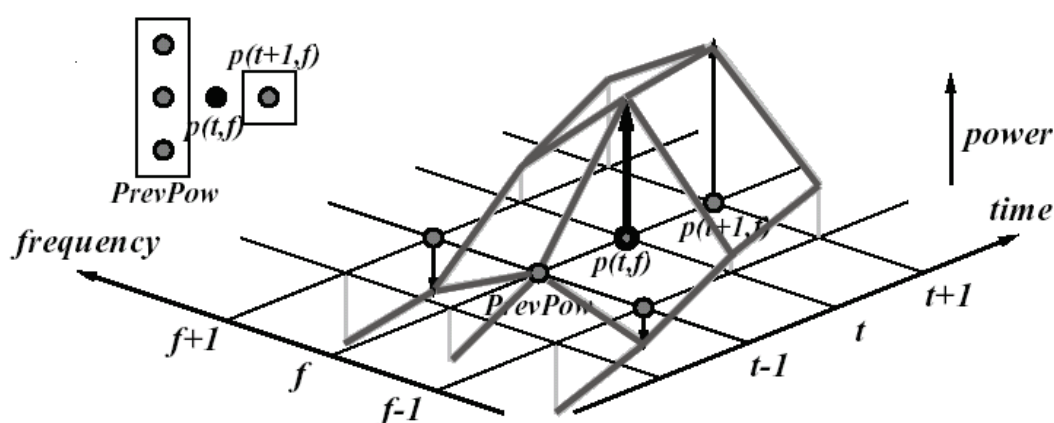


図 3.1: 発音時刻の検出

式 3.1 では時刻 t と時刻 $t+1$ における 2 つのスペクトルパワーを考慮したものになっている. これは発生した音が時間軸方向に連続的に続くという特性を考慮したためである. また, 式 3.2 の算出には対象となっている周波数の上下を考慮している. これはある音が発せられている間, 周波数が微小ながらも変動してしまうことを考慮していることからくるものである.

次に, ユーザが選んだ周波数帯に対して以下に定義する $D(t)$ を算出する.

$$D(t) = \sum_{\text{選択された周波数帯}} d(t, f) \quad (3.3)$$

ユーザが選択できる周波数帯は 0~125Hz, 125~250Hz, 250~500Hz, 500~1000Hz, 1~2kHz, 2~4kHz, 4kHz 以上, の 7 つである. この周波数帯の分割は,

音の周波数が2倍になったときに人間は音が1オクターブ上がったと感ずるという特性を利用したものである。本研究では、ユーザがスペクトログラムを見てどの周波数帯がリズムパートを担っているかを判別し、選択するようにした。

3.3 リズムの開始時刻と平均間隔の算出

3.2に述べた方法で算出した $D(t, f)$ を元に、リズムの開始時刻とリズム間隔の平均値を求める方法について説明する。

3.3.1 平均リズム間隔の算出

平均リズム間隔を求めるにあたり、以下のリズムと発音時刻の関係に関するヒューリスティックな知識を適用する。

知識 1 発音時刻の間隔にリズムの間隔が現れやすい。

$D(t)$ の自己相関関数 $R_{DD}(\tau)$ を算出し、極大となる点を求めることでリズム間隔を求めることができる。

自己相関関数 $R_{DD}(\tau)$ は、 $D(t)$ と τ 遅れた $D(t + \tau)$ の2つの時系列波形がどの程度類似しているか、すなわち $D(t)$ の周期性について示すものであり、式3.4で表すことができる。自己相関関数の概要図を図3.3に示す。

$$\begin{aligned} R_{DD}(\tau) &= \frac{1}{R_{DD}(0)} \frac{1}{T} \sum_{t=0}^T D(t)D(t + \tau) \\ &= \frac{\sum_{t=0}^T D(t)D(t + \tau)}{\sum_{t=0}^T D(t)^2} \end{aligned} \quad (3.4)$$

なお、 $D(t)$ は自己相関関数を算出する前にあらかじめガウシアンフィルタを用いて平滑化を行なった。

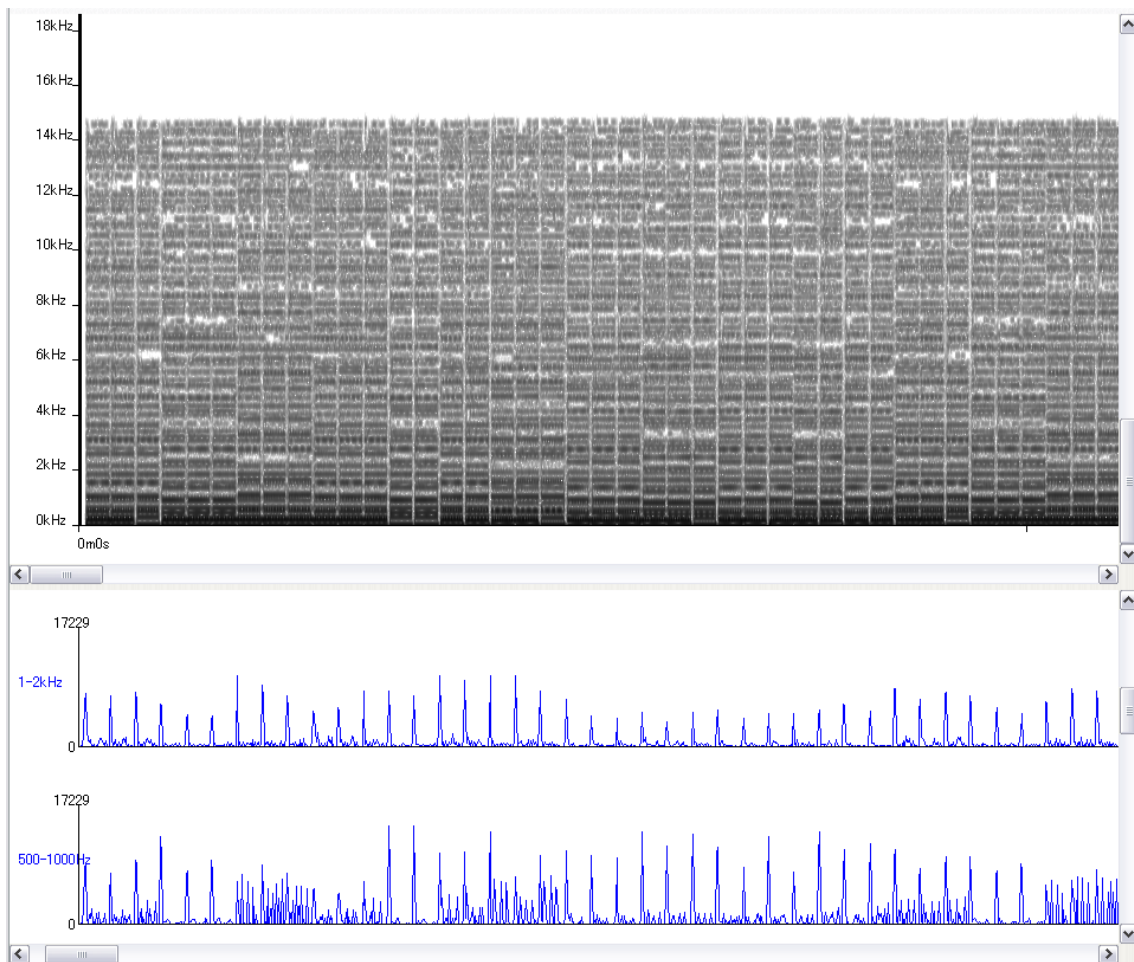


図 3.2: スペクトログラムと各周波数帯発音時刻の例
上段：スペクトログラム
下段：500～1000Hz, および1～2kHz での $D(t)$

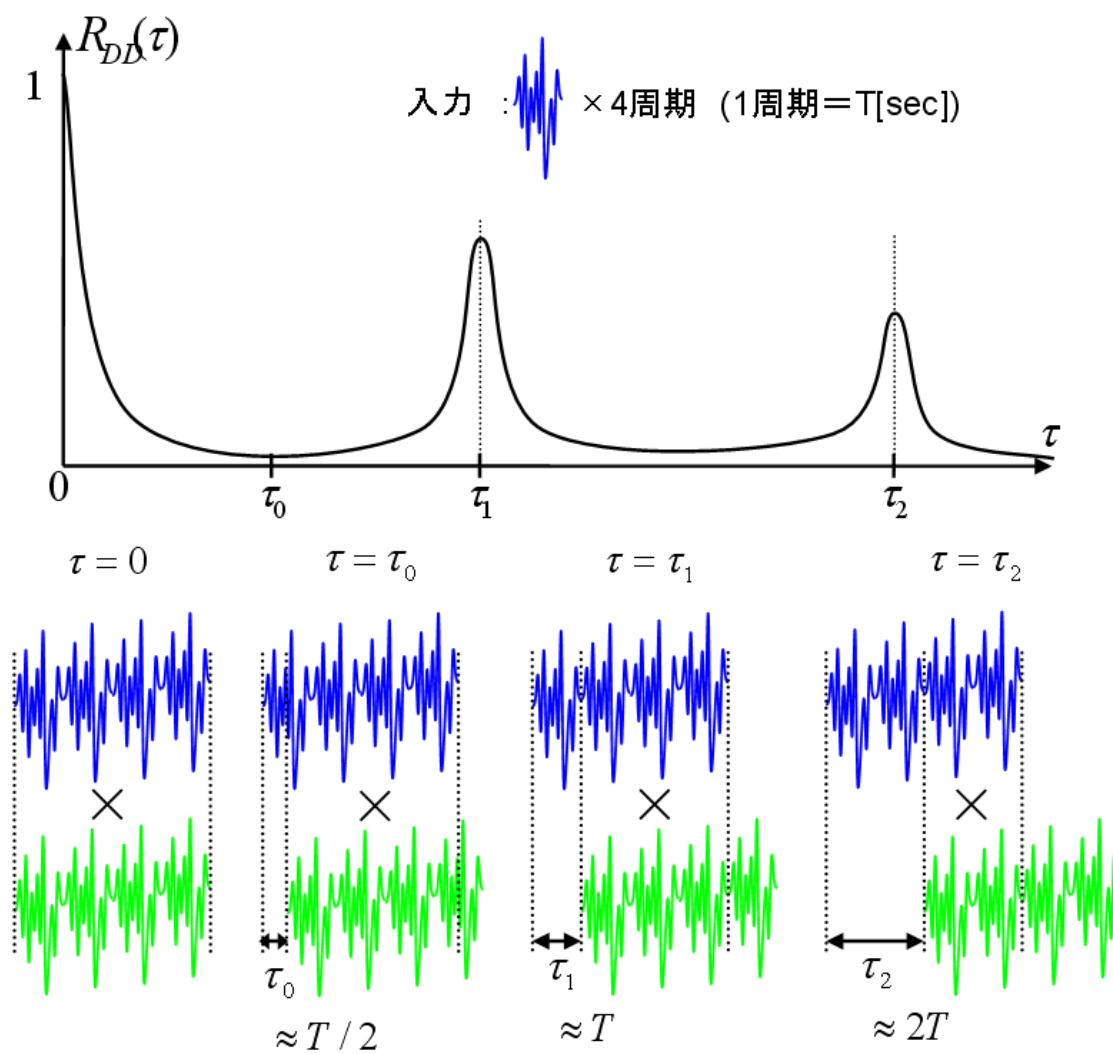


図 3.3: 自己相関関数

3.3.2 リズム開始時刻の算出

リズムの開始時刻の算出において、発音時刻とリズムの関係に関するもう一つのヒューリスティックな知識を利用する。

知識 2 発音時刻の位置がリズムのタイミングである可能性が高い。

3.3.1 で算出されたリズムの間隔と同じ時間間隔のパルス列を作ることができる。このパルス列をリズム時刻の系列とみなし、 $D(t)$ と最もよく一致するようにこのパルス列をずらすことによってビートの開始時刻を算出することができる。リズムのパルス列と $D(t)$ の両者がよく一致する状態は、パルス列と $D(t)$ との相互相関関数を計算し、その極大値によって得られる。

相互相関関数とは自己相関関数と似ていて、2つの時系列波形がどの程度相互依存しているか、もしくは類似しているかを示すもので、2つの波形の位相ずれ時間の関数として表される。今、リズム開始時刻を t_{st} としてリズムのパルス列 $P(t_{st})$ と表し、3.3.1 で算出したリズムの間隔の平均を t_{rhythm} とすると、リズムを示すパルス列は式 3.5 のようになる。

$$P(t_{st}) = \sum_{i=0}^N \delta(t - (t_{st} + i * t_{rhythm})) \quad (3.5)$$

ここで $\delta(t)$ は式 3.6 で表される。

$$\delta(t) = \begin{cases} 1 & (t = 0 \text{ のとき}) \\ 0 & (t \neq 0 \text{ のとき}) \end{cases} \quad (3.6)$$

また N は t_{st} から最後までリズムパルスの数を示す。

よってリズムのパルス列 P と $D(t)$ の相互相関関数 $R_{DP}(\tau_{st})$ は式 3.7 に表すことができる。

$$\begin{aligned} R_{DP}(\tau_{st}) &= \frac{1}{T} \sum_{t=0}^T (D(t) * P(\tau_{st})) \\ &= \frac{1}{T} \sum_{t=0}^T (D(t) * (\sum_{i=0}^N \delta(t - (\tau_{st} + i * t_{rhythm})))) \end{aligned} \quad (3.7)$$

この関数の極大点における τ_{st} の値がリズムの開始時刻となる。

3.4 ビートトラッキング

3.3 ではリズムの開始時刻と平均リズム間隔を求めた。しかし音楽は人の手によって演奏されているため、わずかながらもリズム間隔に誤差が生じる。実際に 3.3 でもとめたリズム間隔で固定してビートトラッキングを行なうと、次第にずれていき、 π 位相誤り (リズムの間隔の半分ずれてしまう誤り) に陥ってしまうような場面が非常に多い。そこで、本研究ではこの問題に対処すべく、以下の手法を用いた。その概要図を図 3.4 に示す。

1. あるリズム時刻 t_i に対して $t_i + t_{\text{rhythm}}$ の前後で $D(t)$ が最大となっている点を探す。
2. 1 で求めた極大点の時刻をリズム時刻 t_{i+1} とする。

この処理を $t_0 = t_{\text{st}}$ (t_{st} : 3.3 で求めたリズム開始時刻) から曲の最後まで繰り返し行なうことで、リズムの揺らぎにも対処できるビートトラッキングができる。

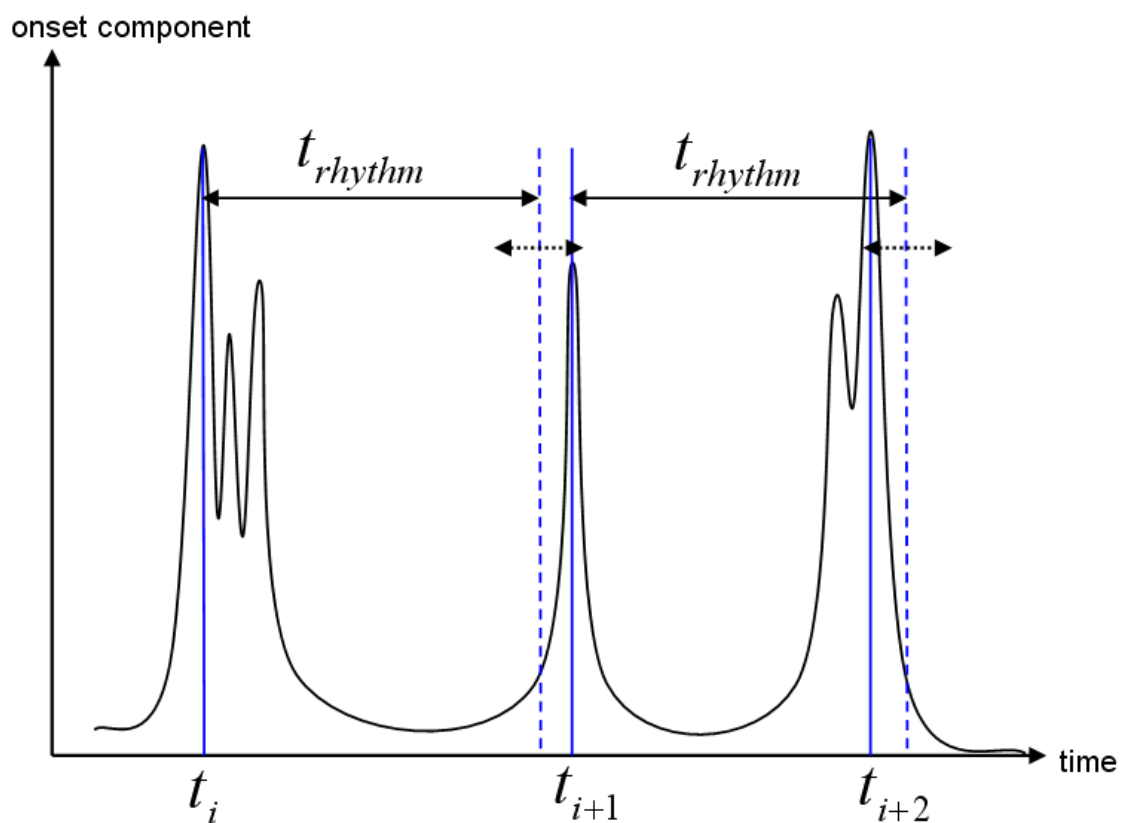


図 3.4: リズムのひずみに対応したビートトラッキング

- 黒の実線：スペクトルパワーの増分
- 青の点線： $t_i + t_{rhythm}$ であらわされる時刻
- 青の実線：本手法で推定されたリズム

第4章

動作プリミティブの抽出

本章では3でもとめた音楽のリズムを元にした、動作プリミティブの抽出方法について述べる。

4.1 動作プリミティブとは

Flash らによると、「人間はすべての動作系列を記憶しているのではなく、有限個の基本動作を記憶しており、その基本動作を組み合わせることによってあらゆる動作を実現している」としている [25]。この基本動作を「動作プリミティブ」と呼ぶ。この動作プリミティブの概念を用いることで、人の動作をシンボリックな記述で表現することが容易になると想像できる。また Flash らは、動作プリミティブの抽出は手や足などのエンドエフェクターの速度変化点で分割することによって行なわれるとしている。すなわち速度の極小点が動作プリミティブの区切りとなる可能性が高い。

舞踊動作においては全身が静止している状態、いわゆる「とめ動作」が存在する。これらは踊り手にとっても舞踊動作におけるキーポーズであると認識されている。よってとめ動作は動作プリミティブを求める上でのキーポーズである可能性が高いということがわかる。またこれに加えて、舞踊は音楽に合わせて踊られる、という特徴を持つ。すなわち、キーポーズは音楽のリズムに基づいたあるリズムパターンに応じて現れるということが推測される。以上の知見から、キーポーズの抽出に関しては以下の条件を用いる。

条件 1 キーポーズでは全身がほぼ静止している

条件 2 キーポーズは音楽のリズムパターンに基づいて現れる

以下ではこの条件が当てはまるキーポーズの算出方法について説明をする。

4.2 解析の対象となる動き

4.2.1 解析の対象となる身体の部位

今回キーポーズを求めるための解析の対象となる動きはエンドエフェクターである両手・両足と重心である。

4.1で述べたように、舞踊動作におけるキーポーズは全身がほぼ静止した状態である。解析の対象として両手・両足を用いるのはこの静止した状態の検出が容易になるためである。

しかし実際の舞踊動作では以下のようなことが起こる可能性が高い。

- 踊りがリズムとずれてしまうことがある
 - － 踊りが音楽のリズムとずれてしまった場合、踊り手は何とか音楽に合わせようとするため、途中の動作が大雑把になってしまうことが多くある。今回対象としている舞踊動作においても同様の傾向が見られ、とめ動作が大雑把になってしまい、しっかりと手先を止めることができなくなっている場面がいくつか見られる。そのためエンドエフェクターの速さだけを用いる方法では、本来とめ動作となるべきところが検出できない可能性がでてきてしまう。
- 個人や性別によって、同じ踊りでも表現がちがうことがある
 - － 伝統舞踊の継承方法は、文書に記録するといった形はとられておらず、人から人への伝承のみに頼っていた。そのため時代がたつにつれてひとつの踊りから様々な踊りに変化していったり、「男踊り」や「女踊り」といった性別による差が生じてきた。また個人によって解釈の差があることから、個人差なども存在する。こういった微小な違いによって、とめ動作のタイミングが異なったり、とめ動作となるべきところでしっかりと前進の動きを停止させないケースが生じる。

こういったケースに対処し、エンドエフェクターのみでは検出できないとめ動作を補うことを目的として、身体の重心を利用した。これは重心が身体の各部位の質量で

重み付けをした平均であるため、とめ動作となるべきところでエンドエフェクターの速度では静止しないように認識される場合でも、全身が静止しているかどうかを判断できるからである。本研究における重心座標の算出には [27] に掲載されているモデルを用いた (図 4.1).

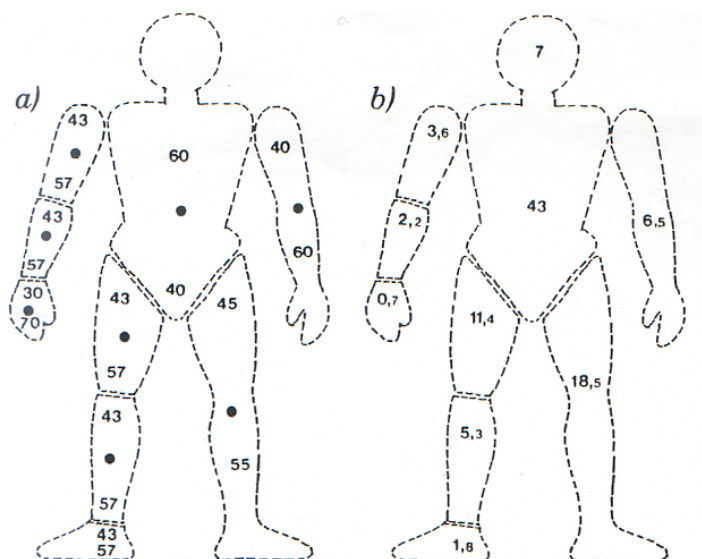


図 4.1: 重心座標のモデル

a) : 身体の各部位の重心位置

b) : 全体重を 100 としたときの身体の各部位の相対重量

4.2.2 解析に用いる座標系

本研究における手の動きの解析においては、各フレームごとに体中心座標系と呼ばれる座標系を定義して座標変換を行い、その体中心座標系における速度を求めて解析を行なった。体中心座標系とは、人間が直立している状態で、腰の位置を原点とし、右肩から左肩にかけての方向を x 軸、人間の正面の方向を y 軸、 x 軸と y 軸に垂直で腰から頭にかけての方向を z 軸と定義するものである (図 4.2).

一方、足および重心の解析においては、データが記録されているままの値 (世界座標系) を用いた。これは足の場合は地面に接地していることが多く、接地している間は世界座標系における速度がほぼ 0 になっている、すなわち足が動いているとき

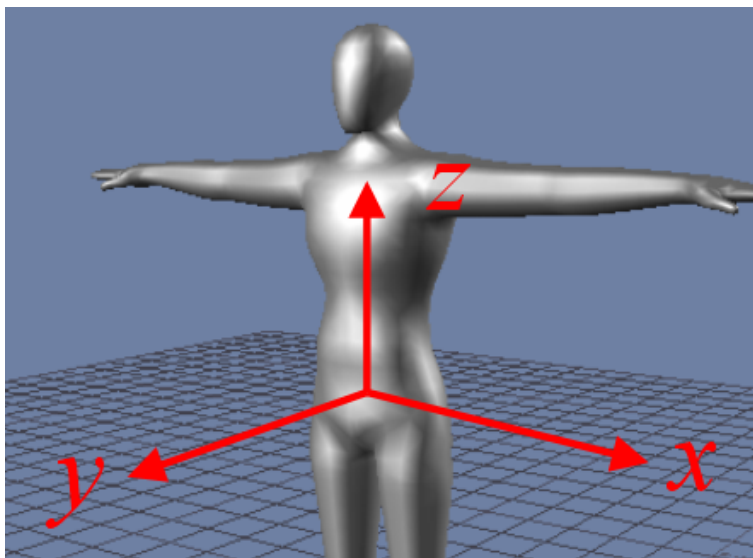
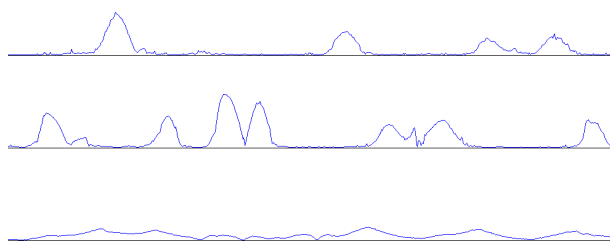


図 4.2: 体中心座標系

と動いていないときの認識が行いやすいという利点に由来するものである．また重心の場合も同じように，とめ動作のときは身体全体が静止している．すなわち世界座標系における重心の速度がほぼ0になっており，身体が動いているときと動いていないときの認識が行いやすいという利点から，世界座標系の値を用いて重心の速度算出を行った．

図 4.3: 足や重心のスピード
(上から左足，右足，重心のスピード)

なお，速度は位置の微分によって求められる．微分は非常にノイズの影響を受けやすいという欠点があり，実際に数値をそのまま微分しただけでは計測時に生じた

ノイズの影響が顕著に現れる。このことを考慮して、まず解析の対象となる手・足・重心の位置をガウシアンフィルタで平滑化し、その後に微分をして速度を求めた。

4.3 セグメンテーション候補点の算出

次に4.2で算出した体の各部位の速さを用いてセグメンテーションの候補点を求める。セグメンテーションの候補点の算出方法は手・重心と足とで異なる。まず最初に手・重心の場合のセグメンテーション候補点の算出方法について述べる。

4.3.1 手・重心の場合

舞踊動作も含めたほとんどの動作シーケンスの中で、手や重心の速さはとめ動作などで一瞬速さがほぼ0になるが、速さ0の状態がしばらく継続することはほとんどない。この特徴を用いて、手・重心におけるセグメンテーション候補点の算出は図4.4に示すフローチャートのもとで行われる。

4.3.2 足の場合

足の速さに関しては、4.2で述べたように接地している間は速さが0である、という特徴を持つ。この特長を生かし、以下の方法でセグメンテーションの候補点を算出した(図4.6)。

1. 速さの立ち上がりと立下りを算出する
2. 立ち上がり～立下り間のグラフの面積を算出する
3. 算出した面積に対してしきい値処理を行う
4. しきい値処理を満たす場合、その立ち上がり、立下りの時点をセグメンテーションの候補点とする

この2で算出している面積は、足の1歩分の軌跡に沿った線積分、すなわち足の移動距離を示している。足の移動距離に対してしきい値処理を行うことで、ノイズの除去を行っている。

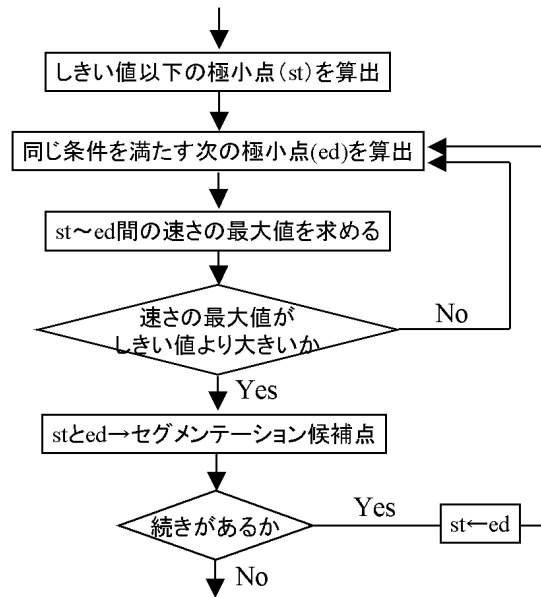


図 4.4: 手・重心のセグメンテーション候補点算出の流れ

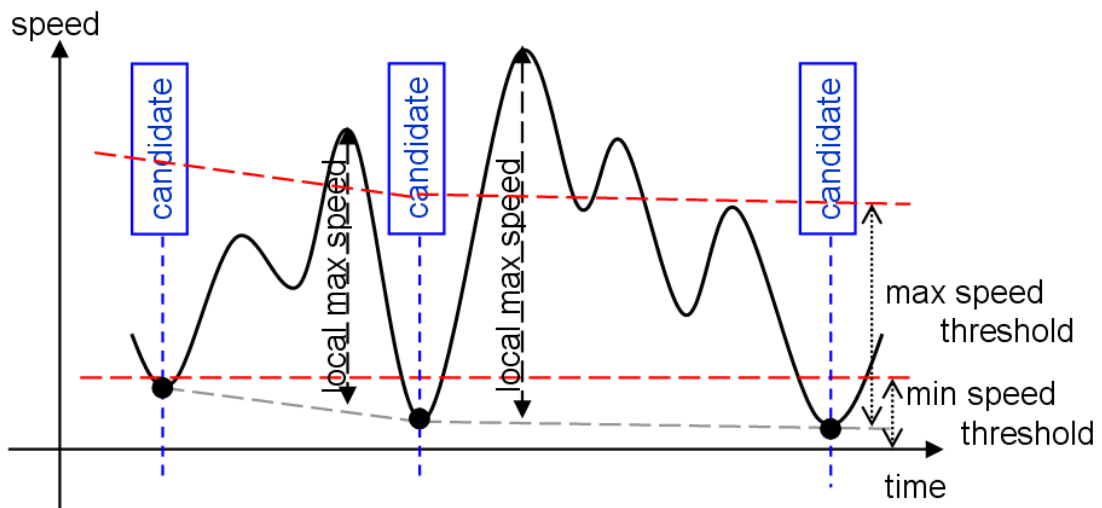


図 4.5: 手・重心のセグメンテーション候補点の算出方法
 条件 1 : $\text{local minimum speed} < \text{minimum speed threshold}$
 条件 2 : $\text{local maximum speed} > \text{maximum speed threshold}$

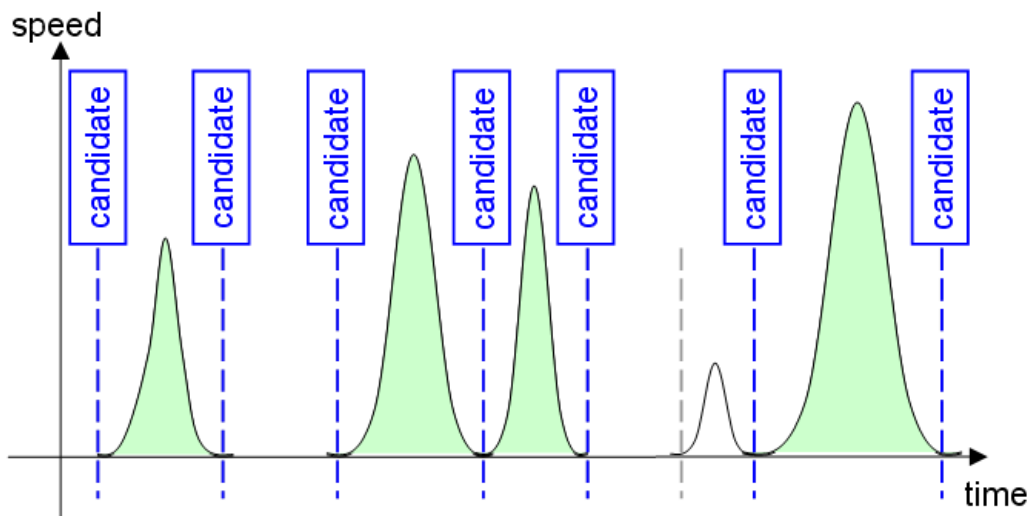


図 4.6: 足のセグメンテーション候補点の算出方法
灰色の破線はしきい値処理によって候補点から除外されたものである

4.4 身体の各部位のセグメンテーション

次に第3章で推定した音楽のリズムを用いて、身体の各部位に関してセグメンテーションを行う。

まず音楽のリズムと動きの同期を取る。本研究ではこの同期は作成したGUIを用いて手作業にて行った。次に音楽のリズムの前後に4.3で求めたセグメンテーション候補点が存在するか確認をする。このときにこの条件を満たす音楽のリズムがあれば、そのリズム時刻でセグメンテーションを行う。

図4.7に身体の各部位のセグメンテーションの例を示す。点線は音楽のリズム時刻、破線はセグメンテーションの候補点を示す。1番目、および3番目のリズムの前後にはセグメンテーションの候補点が存在しないため、この時刻ではセグメンテーションを行わない。一方2番目、および4番目のリズムの前後にはセグメンテーションの候補点が存在している。よってこの時刻においてこの部位の動作シーケンスはセグメンテーションされることになる。

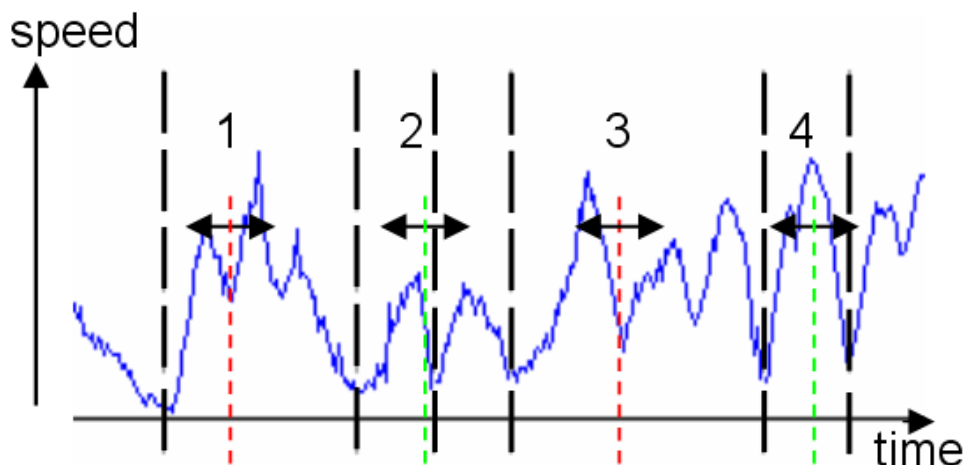


図 4.7: 身体の各部位のセグメンテーション

破線：セグメンテーション候補点

点線：音楽のリズム (このうち緑の点線はセグメンテーション結果)

4.5 動作プリミティブの抽出

4.4 の結果を踏まえて、セグメンテーションを行い動作プリミティブを抽出する。ここでのセグメンテーションは、4.4 の結果を以下に示す論理式に当てはめて行う。

重心の結果 AND (左手, 右手, 両足の結果のうち
2つ以上が一致している場合)

この条件式では「左手, 右手, 両足の結果のうち2つ以上が一致」としている。まず左手, 右手, 両足の結果すべてが一致している場合だが、このときはとめ動作ではないところ抽出してしまうことがある。たとえば歩行動作のように、両足が接地している瞬間では両腕が前、または後ろに伸びていて手が身体に対して静止した状態になっているが、身体全体では前進している。舞踊動作においてもこのような状態が存在しており、エンドエフェクターが静止しているかどうかを認識するだけではとめ動作であるかどうかを認識できない。そこで重心の結果との AND をとることでこの問題を解決している。

また，左手，右手，両足の結果のうち 2 つが一致している場合では，4.2.1 に挙げたようなケースが起こっていて，とめ動作であるはずがとめ動作と認識されない可能性がある．そのために重心の結果を考慮することで，とめ動作の認識の精度を上げている．

第5章

実験

本章では，本研究の提案手法の評価を行う．

5.1 実験データ

今回実験の対象となるデータは福島県会津地方に伝わる会津磐梯山踊りと呼ばれる舞踊の男性と女性 (図 5.2)，およびじょんがら節 (図 5.3) と呼ばれる青森県津軽地方に伝わる舞踊である．

会津磐梯山踊り，じょんがら節ともに，動きデータは会津大学産学イノベーションセンターに設置されている Vicon 社 [28] 製の Vicon Motion Systems を用いて取得した．このモーションキャプチャシステムは光学式モーションキャプチャシステムで，32 個のマーカ (図 5.1) の動きを秒間 120 フレームで測定する．足は片方につき 4 つのマーカがついているが，このうちつま先とくるぶしのマーカのみを解析に使用し，その他 2 つは解析に用いなかった．

一方音楽情報に関しては，実際に舞踊動作を測定したときに使われたカセットテープから USB オーディオデバイスを経由して PC に取り込み，wav 形式に変換したものを利用した．wav ファイルのパラメータについて表 5.1 に記す．

表 5.1: wav ファイルのパラメータ

チャンネル	ステレオ
サンプリングレート	32000 [Hz]
データ速度	128000 [Byte/sec]
ブロックサイズ	4 [ch · Byte/samp]
量子化数	16 [bit/samp]

5.2 ビートトラッキングの結果

リズム推定の際は，フーリエ変換を用いて音響信号を時間領域から周波数領域に変換する．フーリエ変換の窓長は 1024 サンプル，フーリエ変換 1 回につき窓関数を

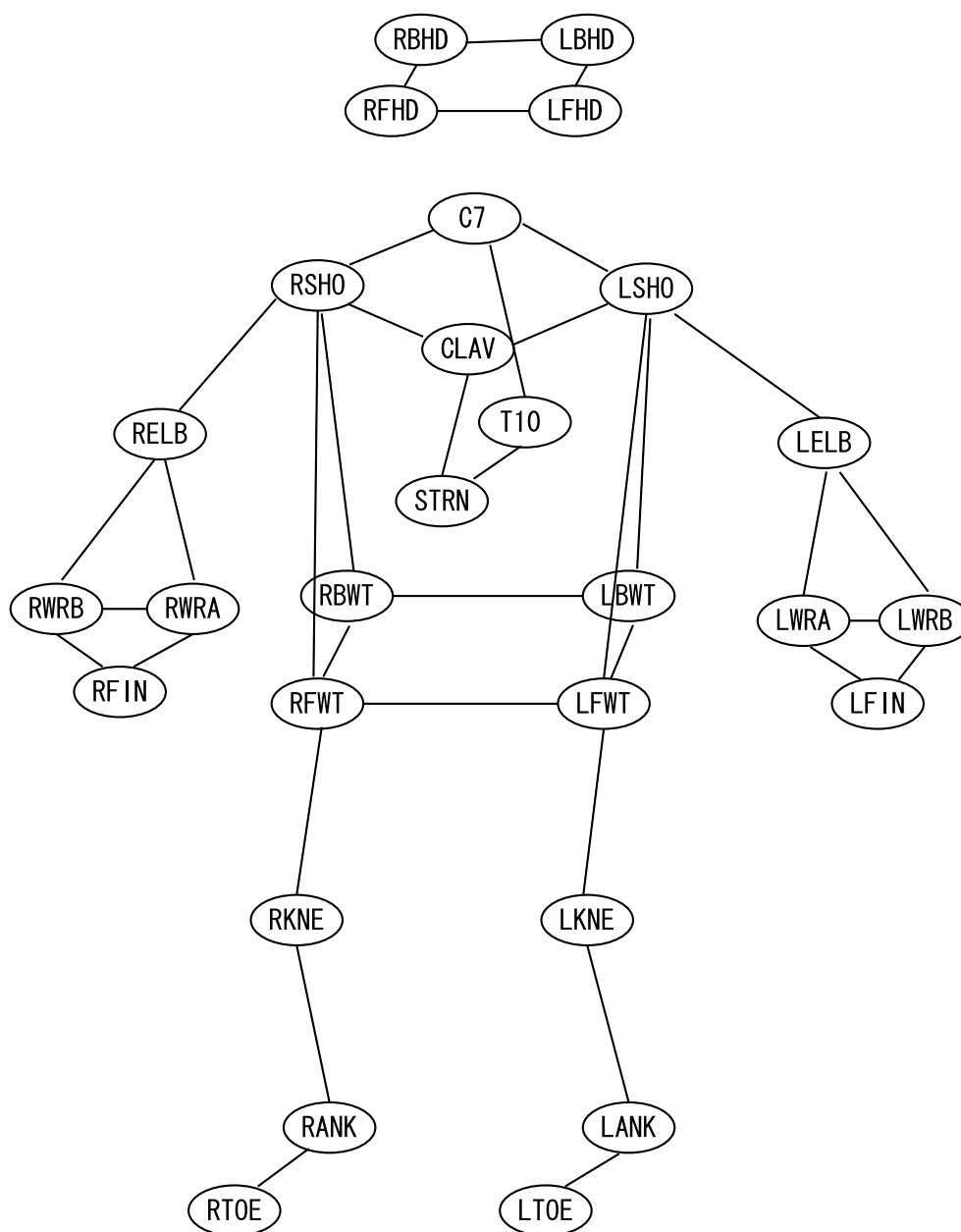


図 5.1: Vicon Motion Systems で使われるマーカー



図 5.2: 会津磐梯山踊り

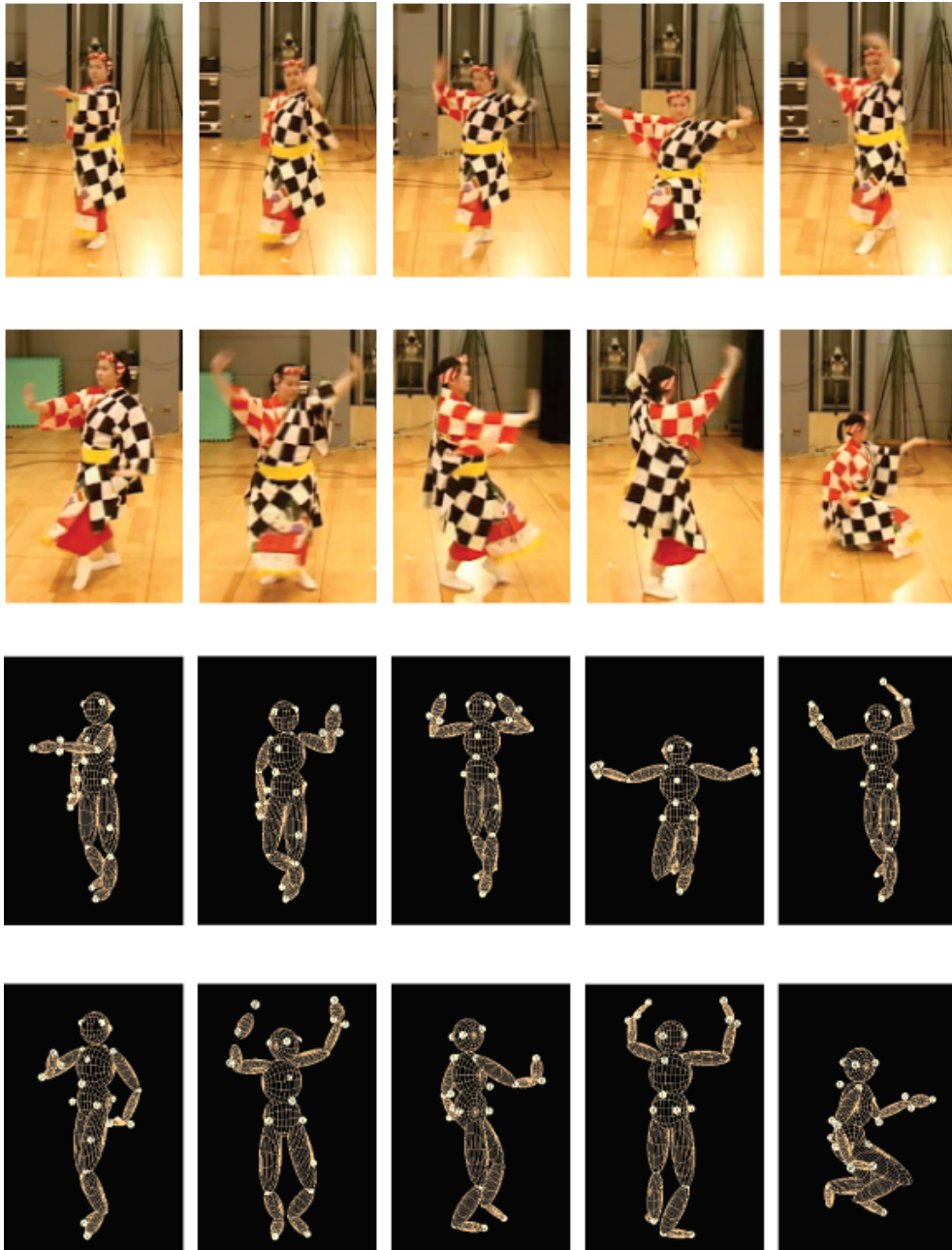


図 5.3: じょんがら節

256 サンプルずつシフトさせていった。また窓関数としては音楽情報処理に向いているといわれているガウス関数を用いた [29]。

5.2.1 リズム間隔推定の誤差評価

本研究で提案したビートトラッキング手法では自己相関関数を用いてリズムの間隔を推定している。この推定されたリズム間隔の誤差についてまず評価を行なう。評価方法としては、ユーザが与えた間隔でビーブ音が記録されているデータを作成し、そのデータに対してリズム間隔を推定し、誤差を求めた。その結果を表 5.2 に示す。

表 5.2: 平均リズム間隔の推定値とその誤差

真値 [sec]	0.5	0.55	0.6	0.65	0.7	0.75
推定値 [sec]	0.48	0.528	0.576	0.624	0.68	0.728
誤差 [%]	4.0	4.0	4.0	6.3	2.9	2.9
真値 [sec]	0.8	0.85	0.9	0.95	1.0	
推定値 [sec]	0.776	0.824	0.88	0.928	0.976	
誤差 [%]	3.0	3.1	2.2	2.3	2.4	

誤差は約 2%~4%に集中していることがわかる。この誤差の原因は以下の 2 つの事象によるものと思われる。

要因 1 スペクトルパワーの時間変化は、短時間フーリエ変換における窓関数をシフトさせていくことで求められる。この窓関数のシフト数を ShiftSamp 、またデータのサンプリング周波数を $\text{SampRate}[\text{Hz}]$ とすると、 $\text{ShiftSamp}/\text{SampRate}[\text{sec}]$ 単位でのスペクトルパワーを把握することができることになり、リズム間隔の推定における最小精度は $\text{ShiftSamp}/\text{SampRate}[\text{sec}]$ ということになる。本研究の場合、 $\text{ShiftSamp}=256$ 、 $\text{SampRate}=32000\text{Hz}$ であるため、 $0.008[\text{sec}]$ が最小精度となる。

要因 2 ある時刻 t におけるスペクトルを求める際、 $t \sim t + (\text{Samp}_{\text{FFT}}/\text{SampRate})$ がフーリエ変換の対象となる。この範囲の間で音が発音された瞬間があると、ス

ペクトルの中にこの音の周波数成分が含まれ、スペクトルパワーが増加して時刻 t で発音されたと認識されてしまう。これが原因となって、正確なリズム間隔が推定できなくなっていると考えられる。

5.2.2 舞踊音楽のリズム推定結果

会津磐梯山踊りの平均リズム間隔は0.704秒(約85M.M. M.M.とはMälzel's Metronomeの略で、1分間の拍数を表す)、じょんがら節の平均リズム間隔は0.584秒(約102M.M.)という結果が出た。それぞれの音楽のスペクトログラムに、推定されたリズム開始時刻およびリズム間隔を線で書き加えたものを図5.4および図5.5に示す。

会津磐梯山踊りは500~1000Hzの周波数領域にかけてリズムを構成している音がスペクトログラム上に見られるが、その発音時刻のところにリズムを示す線が現れていることがわかる。じょんがら節も同様に、広い周波数帯域に渡ってリズムを構成している音がスペクトログラム上に見られるが、その発音時刻のところにリズムを示す線が現れていることがわかる。

また人によってリズムの抽出を行ったものと比べたところ、本研究で提案するビートトラッキング手法がかなり良い精度でリズムをトラッキングできていることが確認できた。

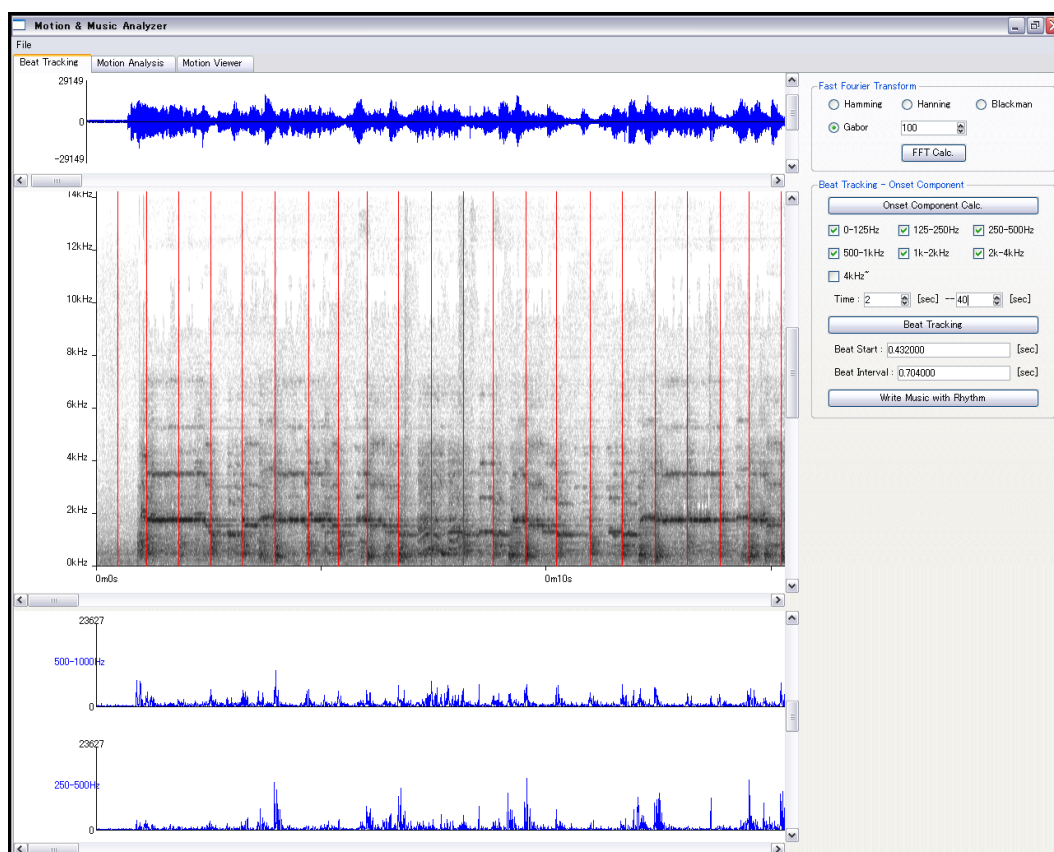


図 5.4: ビートトラッキング結果 - 会津磐梯山踊り
上段: 振幅
中段: スペクトログラム (赤い線は推定されたリズム)
下段: スペクトルパワーの増分 $D(t)$

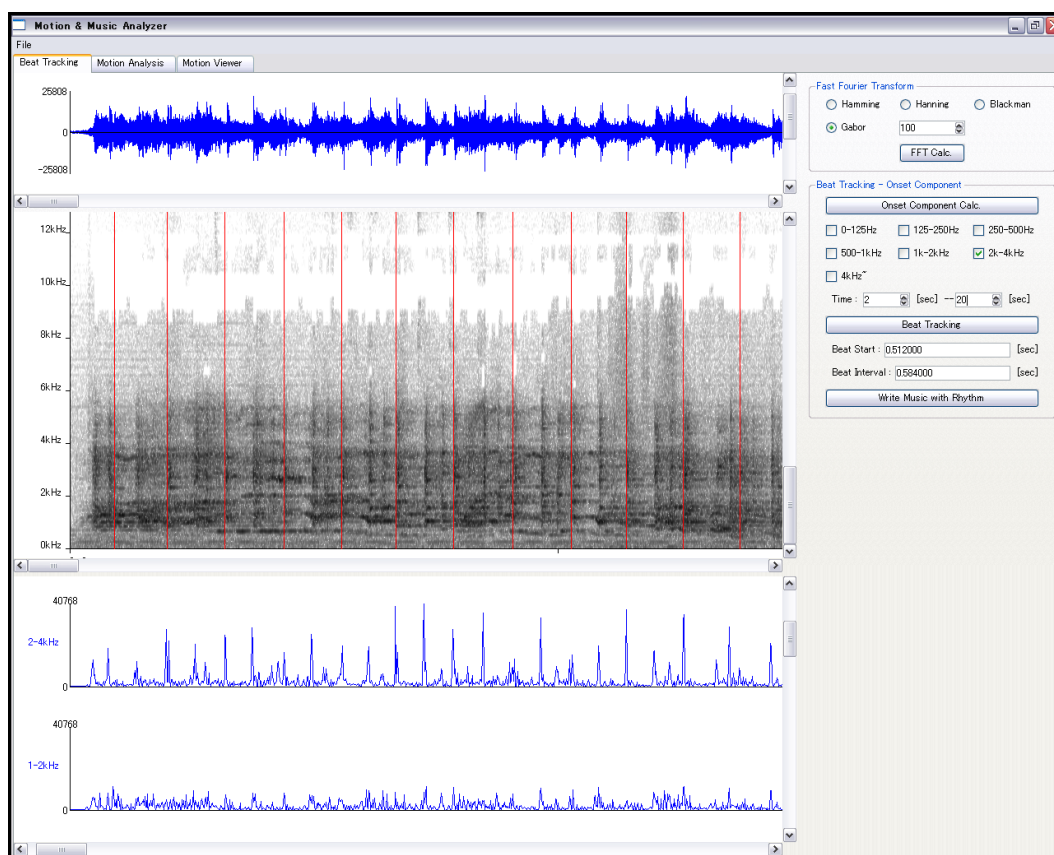


図 5.5: ビートトラッキング結果 - じょんがら節
上段: 振幅
中段: スペクトログラム (赤い線は推定されたリズム)
下段: スペクトルパワーの増分 $D(t)$

5.3 動作プリミティブ抽出の結果

5.2 の音楽のリズム推定結果を用いて動作プリミティブの抽出結果を示す。

また踊り手が実際に踊っているときにどういった姿勢を意識しているか、インタビューを行ない、その結果を元に本研究の評価を行った。会津磐梯山踊りにおける踊り手の意識するキーポーズとその動きの示す意味を表 5.3 に、またじょんがら節における踊り手の意識するキーポーズを表 5.4 に示す。

表 5.3: 踊り手にとっての会津磐梯山踊りのキーポーズ



1 → 2 : 左手をかざして磐梯山を見る



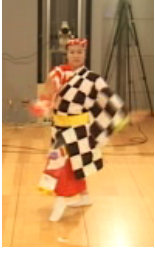
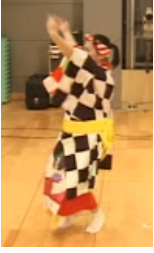


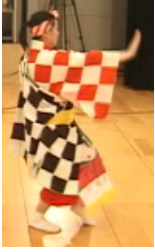


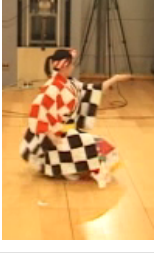


4 : 左手は磐梯山, 右手は猪苗代湖を指す

5 → 6(1) : 磐梯山の形を描く

2 → 3 : 右手をかざして磐梯山を見る

4 → 5 : 稲穂が揺れる様子を示す

表 5.4: 踊り手にとってのじょんがら節のキーポーズ

			
1	2	3	4
			
5	6	7	8
			
9	10	11	12

5.3.1 女性の踊り手による会津磐梯山踊りの場合

対象としたデータは舞踊動作 2 周期分を記録しているものである。女性の踊り手による会津磐梯山踊りでの動作プリミティブ抽出結果を図 5.6 に示す。また比較として、我々の研究グループが以前提案した両手両足のみを用いた動作プリミティブの抽出方法 [13] による結果との比較を表 5.5 に、本手法の結果と表 5.3 との比較を図 5.7 に示す。

表 5.5: 女性の踊り手による会津磐梯山踊りの結果比較

	中澤らの手法 [13]	本研究の提案手法
検出数	13	9
正解数	8	9
未検出数	3	2
誤検出数	5	0

このように女性の踊り手による会津磐梯山踊りにおいては、中澤が提案した手法では誤検出、未検出ともに多く存在するのに対し、本研究の提案手法では誤検出は存在せず、未検出エラーが 2 箇所に見られる程度にとどまった。この未検出エラーは表 5.3 中の 5 番目の姿勢が検出されなかったことを示している。表 5.3 の 5 番目のフレームは、4 → 5 から 5 → 6 へとシームレスに遷移していく中での動作の意味が切り替わる瞬間であり、各エンドエフェクターは動いた状態で現れるキーポーズである。本手法ではエンドエフェクターがほぼ静止しているため動作を検出することを想定しているため、検出できなかったものと見られる。

5.3.2 男性の踊り手による会津磐梯山踊りの場合

対象としたデータは、女性の場合と同様に舞踊動作を 2 周期分記録しているものである。男性の踊り手による会津磐梯山踊りの動作プリミティブ抽出結果を図 5.8 に示す。また、音楽のリズムと両手両足の情報のみを用いて重心は考慮しない場合の結果と比較したものを表 5.6 に、本手法の結果と表 5.3 との比較を図 5.9 に示す。

表 5.6: 男性の踊り手による会津磐梯山踊りの結果比較

	重心を考慮しない場合	本研究の提案手法
検出数	3	7
正解数	3	7
未検出数	8	4
誤検出数	0	0

この男性の踊り手は両手を大きく動かす傾向が大きい。そのためエンドエフェクターの速度が速くなり、とめ動作においてもエンドエフェクターの速度は小さくなっているが、静止していると言えるほど小さくはならない、という特徴がある。

重心を考慮しない場合はほとんどのとめ動作を検出できなかった。これは、とめ動作のところではエンドエフェクターの速度が小さくなってはいるが、静止していると言えるほどの速度にはなっていないため検出できていないと考えられる。この男性は、踊るときにしっかりととめ動作を示すのではなく、あまり強調しないためこのようなことが起きていると考えられる。

また、提案手法では未検出のとめ動作が4つ見られる。このうち2つは表 5.3 中の4が検出されなかったことを示している。これは抽出されたリズムに対して半拍(約0.352秒)早く手を動かしてしまい、リズムに合っていないために検出されなかったと考えられる。残りの2つは表 5.3 中の5が検出されなかったことを示しており、これは女性の踊り手による会津磐梯山踊りの場合と同じ理由によるものと考えられる。いずれの誤差も踊り手が速くエンドエフェクターを動かしていることによるものであることがわかる。

5.3.3 じょんがら節の場合

じょんがら節においては、我々の研究グループが以前提案した両手両足のみを用いた動作プリミティブの抽出方法 [13] と本研究の提案手法とを比較する。それぞれのとめ動作抽出の結果を表 5.7 に示す。また、本手法の結果と表 5.4 との比較の一部を図 5.7 に示す。

表 5.7: じょんがら節の結果比較

	中澤らの手法 [13]	本研究の提案手法
検出数	6	9
正解数	6	9
未検出数	6	3
誤検出数	0	0

じょんがら節における動作プリミティブ抽出の結果を図 5.10 および図 5.11 に示す。

以前提案した抽出方法ではとめ動作と認識できたのは 6 箇所であった。抽出された姿勢はじょんがら節の特徴を示しているキーポーズであるが、検出率は 50% となってしまう。これは、じょんがら節の音楽のリズムが速く、男性の会津磐梯山踊りのようにエンドエフェクターを速く動かしているため、各エンドエフェクターの静止点がタイミングよく音楽のリズムと合うことが難しいため、と考えられる。

一方、我々の提案手法では、3 箇所の未検出にとどまり、検出率は 75% となった。これも前述の理由とほぼおなじで、エンドエフェクターを速く動かす舞踊動作であるため、音楽のリズムのタイミングですべてのエンドエフェクターが静止することが難しいためと考えられる。

以上の結果から、動作プリミティブ抽出における誤差を過去に提案された手法よりも小さくできていることがわかる。また、抽出されたキーポーズは踊り手にとってのキーポーズとほぼ一致しており、手作業ではなく自動で舞踊譜を作成する方法への手がかりになっているとも言える。

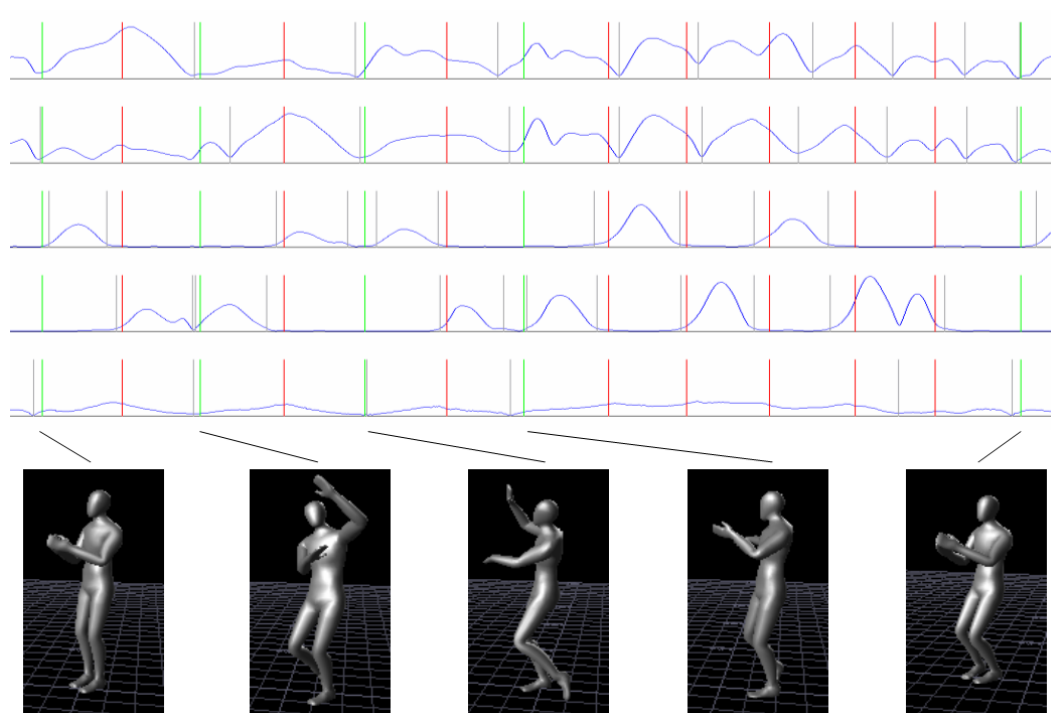


図 5.6: 女性による会津磐梯山踊りのキーポーズ抽出結果
グラフの上から左手, 右手, 左足, 右足, 重心の速さ
灰色の実線: セグメンテーションの候補点
赤の実線: 音楽のリズム
緑の実線: 求めたキーポーズ

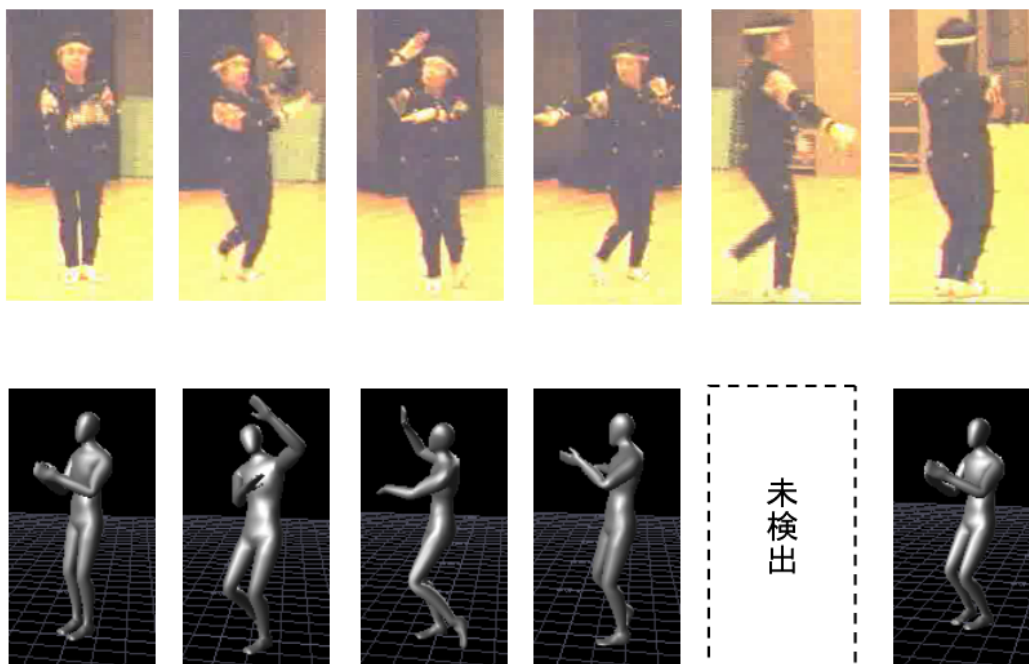


図 5.7: 女性による会津磐梯山踊りの結果の比較

上段：踊り手にとっての動作理解

下段：本手法によるキーポーズの抽出結果

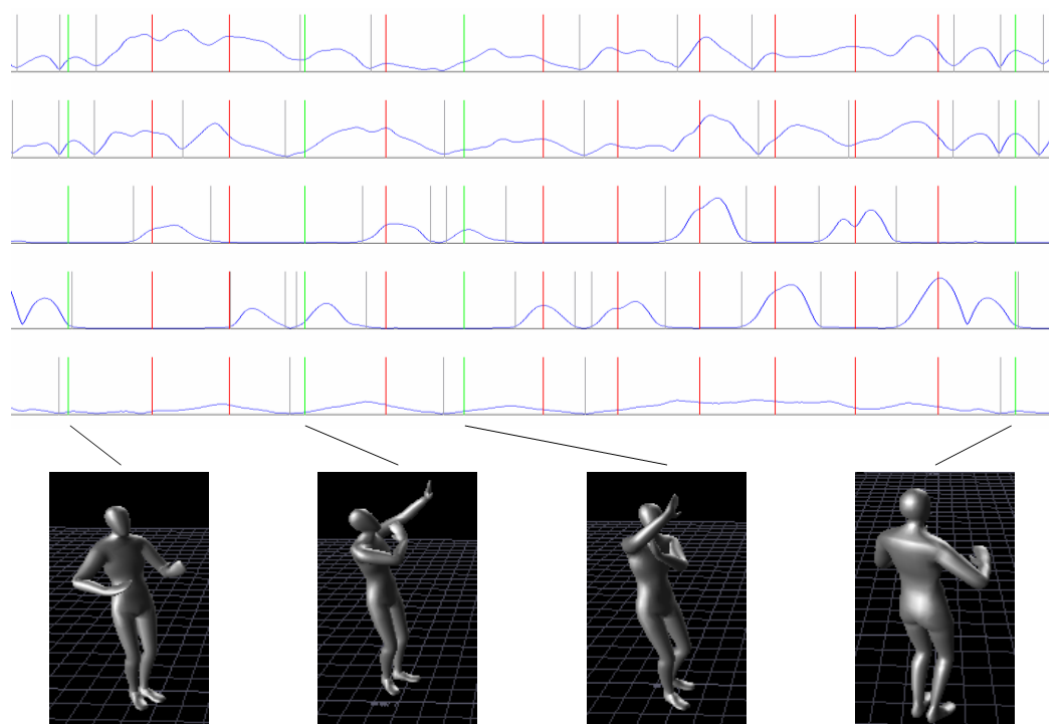


図 5.8: 男性による会津磐梯山踊りのキーポーズ抽出結果
グラフの上から左手, 右手, 左足, 右足, 重心の速さ
灰色の実線: セグメンテーションの候補点
赤の実線: 音楽のリズム
緑の実線: 求めたキーポーズ

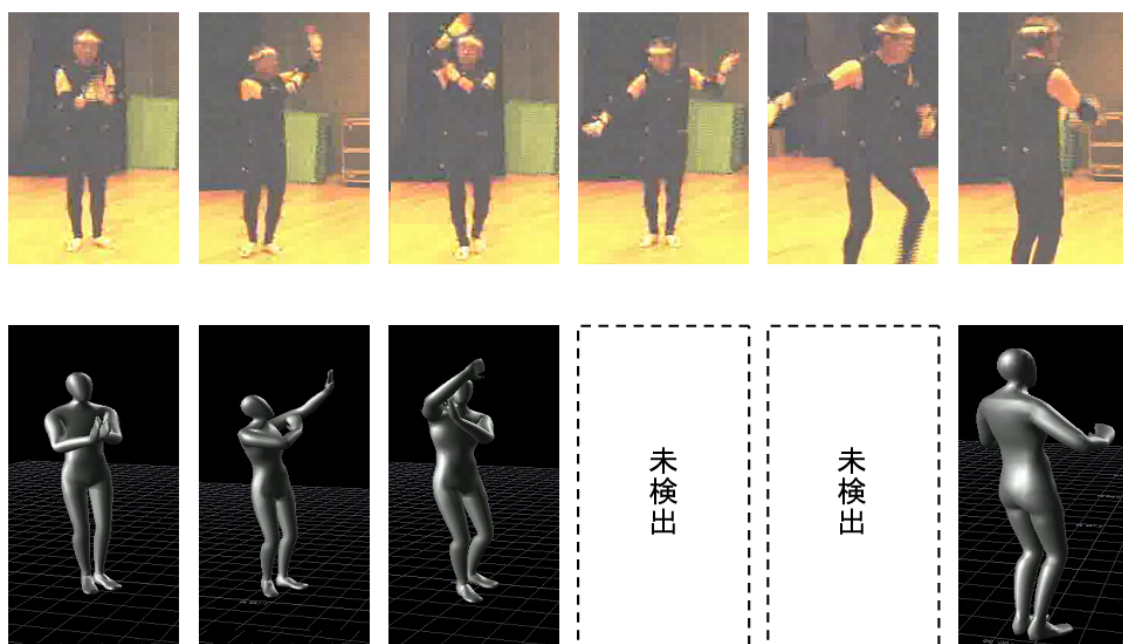


図 5.9: 男性による会津磐梯山踊りの結果の比較
上段：踊り手にとっての動作理解
下段：本手法によるキーポーズの抽出結果

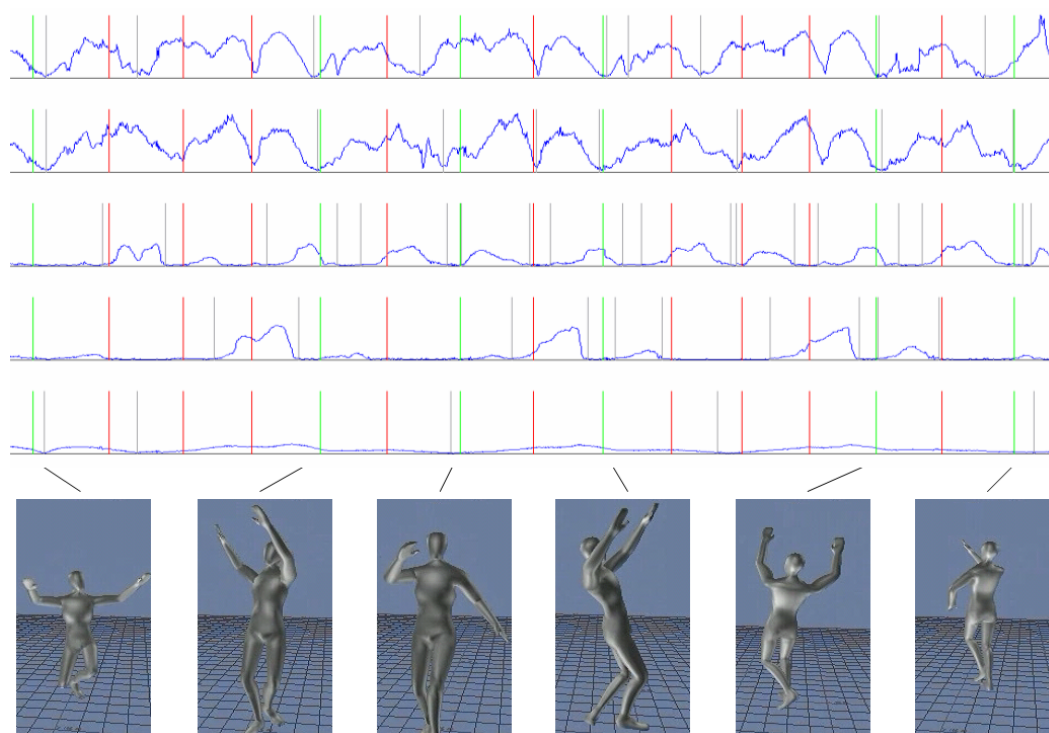


図 5.10: じょんがら節のキーポーズ抽出結果 (一部)
グラフの上から左手, 右手, 左足, 右足, 重心の速さ
灰色の実線: セグメンテーションの候補点
赤の実線: 音楽のリズム
緑の実線: 求めたキーポーズ



図 5.11: じょんがら節のキーポーズ抽出結果 (全て)

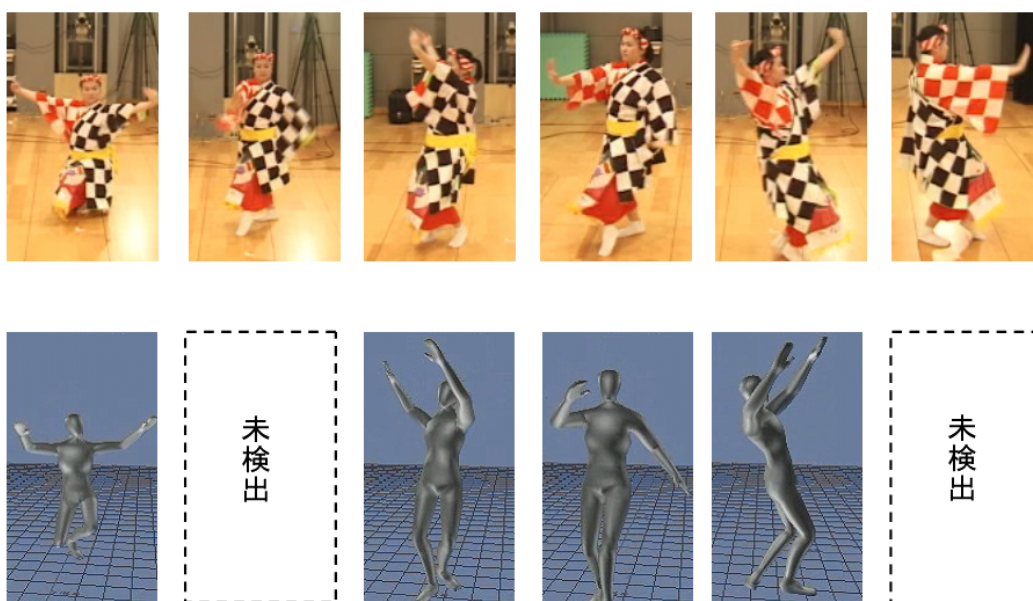


図 5.12: じょんがら節の結果の比較 (一部)

上段: 踊り手にとっての動作理解

下段: 本手法によるキーポーズの抽出結果

第6章

結論

本研究では舞踊動作におけるキーポーズ＝とめ動作の抽出を目的とする解析手法を提案した。

本研究では舞踊は音楽と密接な関係があることに注目し、音楽情報を解析し、その結果を考慮して動作解析を行なう手法を提案した。具体的な解析手法は以下の通りである。

- 舞踊に使われる音楽のビートトラッキング
- 推定されたリズムを考慮したキーポーズの抽出

舞踊音楽は人によって演奏されるため、微小ながらもリズムの揺らぎが多く存在する。本研究での提案手法では、この微小なリズムの揺らぎに対してロバストなビートトラッキングの手法を提案し、実験を行なった。その結果から、精度良くリズムが推定できていることがわかる。

そして、推定された音楽のリズムを考慮した舞踊動作の解析手法を提案した。舞踊動作は個人差などが顕著に現れたり、人間の動きがリズムに合わない場面も多く存在することを考慮して、エンドエフェクターだけでなく重心の速度も考慮にいった解析手法を提案し、実験を行なった。結果は、こういった要因に対してロバストに舞踊中のキーポーズを抽出できることがわかった。言い換えれば、音楽と踊りの間に密接な関係があることを数値解析によって示したともいえる。また、本手法によって抽出されたキーポーズは踊り手が意識しているキーポーズと良い精度で一致しており、新たな舞踊の記述方法への手がかりになると言えるだろう。

本研究の今後の課題として、パラメータの設定や動きと音楽との同期を自動で行なうといったことが挙げられる。現段階ではこれらの設定は、作成した GUI を用いて人間の主観により判断し、解析を行なってきた。これを自動化することによって人間の舞踊における動作理解がより明確になると考えられる。

また本研究の発展として、ロボットが音楽を聴いて舞踊を踊る、すなわち“サウンドフィールドバック制御”が可能になると考えられる。しかし実際にロボットに踊らせるにあたっては、関節角やその角速度に制限があり、動きデータをそのままロボットに適用して踊らせることはできない。そこでまず当面の目標としては、その関節角の制限内で人間らしい動きを最大限に再現できる手法を提案する必要があると考えられる。

これを実現した上で、流れてくる音楽からリアルタイムでリズムを認識し、そのリズムの範囲内で可能な限り人間に近いロボットの動きを生成し、そして踊らせる

ことでサウンドフィードバック機能を実現していきたい。

謝辞

本研究を進めるにあたり、忙しい御身にも関わらず、毎週ミーティングを行なっていたり、的確なアドバイスをしてくださり、またすばらしい研究環境を整えてくださった池内克史教授には心からお礼を申し上げます。また研究だけでなく進路の相談にもものっていただき、深い感謝の意を表します。

そして踊りグループの長として君臨し、いつもの的確な指導・アドバイスをしてくださった大阪大学の中澤篤志講師には本当にお世話になりました。この場を借りて深い感謝および敬意を表します。

また、踊りグループのメンバーである高松淳さん、工藤俊亮さん、中岡慎一郎さん、Miti Ruchanurucks さんには、ミーティングや研究が行き詰ったとき、プログラミングがわからないときなどいろいろな場面で適切なアドバイスをいただきました。本当にありがとうございました。そして踊りグループだけでなく池内研究室の皆さんには本当にお世話になりました。

学会出席や事務手続きで大変お世話になった池内研究室の秘書のみなさま、研究に対するご指導の他にも研究生活を進めるにあたりご尽力して頂いた池内研究室の職員みなさまに感謝いたします。

研究だけでなく、様々な悩みや相談にのっていただき、修士2年間で共に過ごした上原康彦さん、猪狩壮文さん、崎田健二さん、原田貴昭さん、廣田祐一郎さんにも感謝いたします。

最後に学生生活をいろんな面から支えてくれた家族と友人たち、飼い犬のダックスフンドに最大限の感謝の気持ちを表します。

平成16年 1月 30日

参考文献

- [1] 東京国立文化財研究所：“標準日本舞踊譜,” (1960).
- [2] A. Hutchinson: “Labanotation,” (1977).
- [3] Y. Yoshida, Y. Matsuoka and K. Hachimura: “In processing of human body movement based on labanotation labanreader and labaneditor,” IPSJ (1998).
- [4] A. Driesse: “Real-time tempo tracking using rules to analyze rhythmic qualities,” International Computer Music Conference, pp. 578–581 (1991).
- [5] R. Rowe: “Interactive music systems,” The MIT Press (1993).
- [6] D. Rosenthal: “Machine rhythm : Computer emulation of human rhythm perception,” Ph.D. thesis, Massachusetts Institute of Technology (1992).
- [7] B. Vercoe: “Perceptually-based music pattern recognition and response,” In Proceedings of 3rd International Conference for the Perception and Cognition of Music, pp. 59–60 (1994).
- [8] 片寄 晴弘：“音楽感性情報処理に関する研究,” 博士論文, 大阪大学 基礎工学部 (1991).
- [9] N. Todd: “The auditory “primal sketch” : A multiscale model of rhythmic grouping,” Journal of New Music Research, **23**, 1, pp. 25–70 (1994).
- [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] A. Bruderlin and L. Williams: “Motion signal processing,” In Proceedings of ACM SIGGRAPH95, pp. 97–104 (1995).
- [12] 大崎, 嶋田, 上原: “速度に基づく切り出しとクラスタリングによる基本動作の抽出,” 人工知能学会誌, **15 卷 8 号**, pp. 878–885 (2000).
- [13] A. Nakazawa, S. Nakaoka, K. Ikeuchi and K. Yokoi: “Imitating human dance motions through motion structure analysis,” In Proceedings of International Conference on Intelligent Robots and Systems, pp. 2539–2544 (2002).
- [14] M. Brand and A. Hertzmann: “Style machines,” In Proceedings of ACM SIGGRAPH2000, Annual Conference Series, pp. 183–192 (2000).
- [15] A. Witkin and Z. Popovic: “Motion warping,” In Proceedings of ACM SIGGRAPH95, Annual Conference Series, pp. 105–108 (1995).
- [16] M. Gleicher: “Retargetting motion to new character,” In Proceedings of ACM SIGGRAPH98, Annual Conference Series (1998).
- [17] J. Lee and S. Y. Shin: “A hierarchical approach to interactive motion editing for human-like figures,” In Proceedings of ACM SIGGRAPH99, Annual Conference Series, pp. 39–48 (1999).
- [18] C. Rose, B. Guenter, B. Bodenheimer and M. F. Cohen: “Efficient motion generation of motion transition using spacetime constraints,” In Proceedings of ACM SIGGRAPH96, Annual Conference Series, pp. 147–154 (1996).
- [19] L. Kovar and M. Gleicher: “Motion graph,” In Proceedings of ACM SIGGRAPH02, Annual Conference Series, pp. 473–482 (2002).
- [20] J. Lee, J. Chai, P. S. A. Reitsma, J. K. Hodgins and N. S. Pollard: “Interactive control of avatars animated with human motion data,” In Proceedings of ACM SIGGRAPH02, Annual Conference Series (2002).
- [21] O. Arikan and D. A. Forsyth: “Interactive motion generation from eamples,” In Proceedings of ACM SIGGRAPH02, Annual Conference Series, pp. 483–490 (2002).

- [22] T. Kim, S. I. Park and S. Y. Shin: “Rhythmic-motion synthesis based on motion-beat analysis,” In Proceedings of ACM SIGGRAPH03, Annual Conference Series (2003).
- [23] A. Nakazawa, S. Nakaoka and K. Ikeuchi: “Synthesize stylistic human motion from examples,” In Proceedings of International Conference on Robotics and Automation (2003).
- [24] S. Nakaoka, A. Nakazawa, K. Yokoi, H. Kimura and K. Ikeuchi: “Generating whole body motions for a biped humanoid robot from captured human dances,” In Proceedings of International Conference on Robotics and Automation (2003).
- [25] T. Flash and H. Hogan: “The coordination of arm movements,” J. Neuroscience, pp. 1688–1703 (1985).
- [26] 長嶋 洋一, 橋本 周治, 平賀 譲, 平田 圭二: “コンピュータと音楽の世界基礎からフロンティアまで,” 共立出版 (1999).
- [27] Rolf Wirhed 著, 金子 公宥, 松本迪子訳: “目でみる動きの解剖学—スポーツにおける運動と身体のメカニズム,” 大修館書店.
- [28] Oxford Metrics, Vicon: “<http://www.oxfordmetrics.co.uk/>,”.
- [29] Curtis Roads 著, 青柳 龍也, 小坂 直敏, 後藤 真孝, 引地 孝文, 平田圭二, 平野 砂峰旅, 堀内 靖雄, 松島俊明訳: “コンピュータ音楽 歴史・テクノロジー・アート,” 東京電機大学出版局 (2001).