

# ダンスをマスターした自身の映像を 先に見ることによるダンス学習支援

土田 修平<sup>1,a)</sup> Mao Haomin<sup>1</sup> 岡本 秀明<sup>2</sup> 鈴木 裕真<sup>2</sup> 金田 麟太郎<sup>2</sup> 堀 隆之<sup>2</sup> 寺田 努<sup>1,b)</sup>  
塚本 昌彦<sup>2,c)</sup>

**概要：**本研究では、プロフェッショナルなダンサーのダンスをマスターして踊る自身の映像を見ながらダンス動作を学習する学習支援手法を提案する。理想的な動作を行う自身の映像を見ることで行動の改善を促す「ビデオセルフモデリング」は、運動技能の学習においても有効とされる。しかし、理想的な動作を行なっている自身の映像を作成するには、多くの時間を割く必要があり手間がかかる。そこで、深層学習による映像生成技術を用いて参照動画中のダンサーの動作と同じ動作を行う学習者自身の映像を自動で生成し、その映像を見ることで学習者の学習を促す。生成した映像がダンス動作の学習に有効かどうかを検証するために参加者 20 名による評価実験を実施した。実験の結果、他者の映像を提示した場合と生成した自己映像を提示した場合とで有意差は見られなかったが、自己映像を提示したグループの方が、自己効力感が有意に低い傾向となった。これら実験結果を踏まえ、映像生成技術を活用したダンス動作の学習支援手法の可能性を示した。

## 1. はじめに

ダンスや武道、スポーツなどのさまざまな身体活動において、新たな運動技能を効率的に習得することは重要である。運動技能を習得するためには、一般的に何時間ものトレーニングと反復練習が必要となる [1]。その中でもダンスは、楽曲のリズムやメロディのタイミングに合わせて比較的複雑な全身の動作を伴うため、習得に時間がかかる。ダンス動作を効率的に習得するために、これまで視覚、触覚、聴覚などさまざまなモダリティを活用した数多くの運動学習支援手法やツールが提案されている [2-4]。特に映像視聴による学習支援は、これまで多数提案されてきた [5]。藤本ら [6] は、熟練者のダンスで踊る自身の映像を用いた教師映像を用いれば学習が早まるという仮説のもと、ダンスを踊る自身の映像を比較的簡単な姿勢をとるだけで生成できる練習支援システムを提案している。

藤本らが提案するシステムのように、理想的な動作を行なっている自分自身を見ることで、動作の改善を促す手法を「ビデオセルフモデリング」と呼ぶ [7]。Diane ら [8] は、トランポリンの学習において、現在の能力よりも高いスキ

ルレベルでトランポリンの演技をしている学習者の映像を映像編集ソフトを用いて作成し提示することで、パフォーマンスが向上したことを確認している。このことから、理想的な動作を行なっている映像を学習者に見せることは、運動技能の習得において有効である。

しかし、そのような理想的な動作を行なっている学習者自身の映像の作成には時間がかかり、ある程度の編集スキルも必要となる。その上、複雑な動作を伴うダンス動作の場合、楽曲に合わせて理想的な動作で踊る学習者の映像を作成することは容易でない。藤本らの手法においても、フレーム単位で学習者の姿勢を撮影する必要がある、教師動画中のダンサーと同じ姿勢をフレーム数分とすることは手間がかかる。また、フレーム単位でダンサーの姿勢と完全にマッチした姿勢を学習者がとることは困難であるため、ある程度の誤差を許容する必要がある。そのため、この手法では最終的な映像の質を高めることに限界がある。ここで Chan ら [9] が提案する深層学習を用いた映像生成技術を活用すれば、ビデオセルフモデリングが適用可能な、高精度・高精細で踊る自身の映像を自動で生成することができ、その自身の映像を用いればダンスのような複雑な動作の学習を早められるのではないかと考えた。

そこで本研究では、ダンスをマスターした自分を先に見ることによる練習支援システムを提案する。自動生成した自己映像がダンス動作の学習に有効かどうか検証する。具

<sup>1</sup> 神戸大学大学院工学研究科

<sup>2</sup> ソフトバンク株式会社

a) tuchida@eedept.kobe-u.ac.jp

b) tutomu@eedept.kobe-u.ac.jp

c) tuka@kobe-u.ac.jp

体的には、Chan ら [9] の手法を参考にモデルを構築、学習者の練習中の様子を 3 分程度撮影した映像を入力とし、プロフェッショナルなダンサーの参照動画から抽出された動作と同じダンス動作を行う学習者自身による自己映像を作成する。学習者は上手いダンスを踊る自身の映像を見ながら練習することで、効率的な動作の習得ができると予想する。実験参加者 20 名の内、参加者自身が踊る自己映像を提示するグループとダンサーが踊る映像を提示するグループに分けて評価実験を実施した。その結果、学習の效果に有意な差は見られなかったが、自己映像の提示、他者映像の提示と比べて自己効力感が有意に低くなる傾向が見られた。アンケートの自由記述から、実際にはできない難しい動作をこなしている自己映像が、自己効力感を低下させた要因と考える。またアンケート結果から、学習するダンス動作の難易度が高くなるほど、生成された映像を見ても自分が踊っているようには感じられない傾向が見られた。以上のことから、将来的には学習者のスキルレベルに合わせて調整された動作を行う自己映像を生成する手法が有効ではないかと考えられる。

## 2. 関連研究

自分が理想的な行動をとっているビデオを見ることで、その行動を改善したりする方法は「ビデオセルフモデリング」と呼ばれており、これまでさまざまな環境下での利用が報告されている [10–12]。Diane ら [8] は、トランポリンの学習において、学習者が現在の能力よりも高いスキルレベルでトランポリンの演技をしているような映像を映像編集で作成し、提示した。その結果、口頭のみ指示と比べて、映像を用いて学習した動作のパフォーマンスが向上していることを確認した。以上のように、ビデオセルフモデリングは運動技能の学習にも有効である。本研究で提案する手法では、自身の練習している様子の映像を入力として、運動学習に有効なビデオセルフモデリングで使用する動画を自動生成する深層学習モデルを構築する。

ビデオセルフモデリングをインタラクティブに適用する手法もいくつか提案されている。Michael ら [13] は、エクササイズバイクを用いた VR エクササイズゲームを提案している。このゲームでは、プレイヤーの過去のプレイをすべて記録し、将来の自分である「ゴースト」と競争できるレースになっている。4 週間に渡る実験の結果、提案手法の方が身体的パフォーマンス、内発的動機、及びフローを向上させることを確認した。本研究では、上記に挙げたようなインタラクティブなビデオセルフモデリングは実現できていないが、将来的には学習者のスキルレベルに合わせて映像を生成し、リアルタイムに提示可能と考えている。本研究はそれら技術の実現に向けた第一歩として、生成した映像がダンス動作の学習にどのような影響を与えるか確認する。

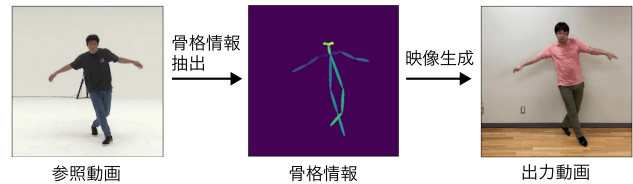


図 1 映像生成フロー

## 3. 見本映像の生成

本研究では、習得するダンス動作の見本となる自己映像の生成に、Chan ら [9] が提案する手法を参考に作成した映像生成モデルを用いた (図 1)。OpenPose [14] を用いて抽出された骨格情報を中間表現として用いて、動画を生成する。骨格情報の正規化により入力動画中の学習者と参照動画中のダンサーの体格やカメラ内位置の差異を吸収する。入力動画は、参照動画に含まれるダンス動作を真似て練習している様子を撮影した動画の 3 分間を用いる。出力動画として、参照動画と同じ長さで参考動画中の人物が行うダンス動作と同じ動作で踊る入力動画中の人物の動画が生成される。学習にはクラウドサービスを使用し、GPU は Tesla V100 (16 GB)、vGPU が 8、メモリが 61GB の環境を用いて、40 エポックで学習させた。すべての処理が終了するまでに、入力動画 1 本あたり約 20 時間かかった。

予備実験として、まず入力動画の検討を行った。自由に動き回った動画を入力した場合と、ダンス動作の練習時と同様に参照動画と同じダンス動作を繰り返した動画を入力した場合とで出力動画を比較した結果、同じ動作を繰り返した動画を入力した場合の方が、質の高い動画が出力されることがわかった。次に、さまざまなダンス動作を含む複数の参照動画とそのダンス動作を真似て練習している様子を撮影した 3 分間の入力動画を用いて学習を行い、動画を出力した。その結果、奥行き方向の動作が、上下方向の動作、ダンサーが後ろを向いていたり、素早く複雑な動作が含まれてのようなダンス動作の場合は、ダンス動作の見本動画としての質が低くなることがわかった。具体的には、奥行き方向動作が含まれる場合、中間表現の骨格情報が 2 次元であるためにうまく奥行き方向の動作を表現できていなかった。上下方向の動作では、動画中のヒトの位置が上下に細く移動してしまった。後ろを向く動作では、後頭部ではなく顔が表示されることがあった。素早く複雑な動作では、腕が突然消失したり、腕の位置がワープするような表現が見られた。以上より、評価実験時にはこれら動作を含まない参照動画を採用する。

## 4. 評価実験

ダンスをマスターした自身の映像を先に見ることがダンス学習に有効かを検証するために評価実験を行った。参加者は 20 代の大学生 20 名 (男性 19 名、女性 1 名) で、ダ

ンスの経験は中学校・高校等での授業、文化祭などの校内イベントで実施された集団でのダンスパフォーマンスに参加した程度である。本実験は、大学内に設置された 43m<sup>2</sup> のフローリングフロアを有するスタジオにて実施した。なお、この実験は神戸大学の研究倫理審査委員会の承認を得て行った。

#### 4.1 学習対象のダンス動作

参加者は三つのダンス動作（図 2 右参照）を学習する。三つのダンス動作は、AIST Dance Video Database [15] の動画からトリミングし、参照動画として使用した。ダンスジャンルはそれぞれ Break, Pop, Pop, 動画の長さは 5.9 秒, 9.6 秒, 8.6 秒, 動画に含まれる楽曲の BPM は 80, 100, 110, 動画の名称は gBR.sBM.c01.d04.mBR0.ch01, gPO.sBM.c01.d10.mPO2.ch10, gPO.sFM.c01.d11.mPO3.ch11 で、論文中では Dance1, Dance2, Dance3 と呼ぶことにする。予備実験を通して得られた知見に基づき、生成される見本映像の質を高めるために、奥行き方向、上下方向の動作が少なく、ダンサーが後ろを向いておらず、素早く複雑な動作が含まれていない、カメラが撮影している方向に対して垂直な平面上を動く動作を中心として構成されたダンス動作を選んだ。

#### 4.2 実験条件

参加者は 10 名の自己映像提示グループと 10 名の他者映像提示グループの二つのグループに分けられる。グループ間の参加者のダンスの熟練度を揃えるために、グループ分けは 4.4 節で説明する PreTest で算出された DTW コストの値に基づき、PreTest での DTW コストの平均値の差が 10 以内に収まるように分けた。尚、自己映像提示グループの DTW コストの平均は 814.5、標準偏差は 61.5、他者映像提示グループの DTW コストの平均は 815.2、標準偏差は 61.9 であった。図 2 左に示すように、左半分は PC 搭載のカメラを通して撮影した参加者の映像を反転させた鏡映像を、右半分には見本映像を表示するアプリケーションを用いて練習する。鏡映像はリアルタイムに表示される。アプリケーションは macOS Big Sur 上の openFrameworks v0.11.2 で開発された。自己映像提示グループでは、練習時（4.3 節の PreTraining）に撮影した映像より生成された、参照動画のダンサーのダンスを踊る自己映像を見ながら練習を実施する。他者映像提示グループでは、ダンサーによる見本映像（参照動画）を見ながら練習を実施する。

三つのダンス動作（Dance1~Dance3）の学習順序は、可能な限り被験者間で被りが少なくなるようバランスをとった。映像は、外部ディスプレイ（90 cm × 50 cm）に表示され、参加者はその映像を見ながら練習を実施する。実験環境を図 3 に示す。カメラは iPhone SE（第 2 世代）、PC は MacBook Pro（13-inch, 2019, Four Thunderbolt 3 ports）

をそれぞれ用いた。PC の画面上には学習の残り時間を示すタイマーを表示した。

#### 4.3 実験手順

実験行程を図 4 に示す。実験は DAY1, DAY2, DAY3 の 3 日間実施された。

DAY1 では、まず参加者はスタジオに入り、これから三つのダンス動作（Dance1~Dance3）をそれぞれ 5 分間ずつ練習することについての説明を受けた。練習時に使用するアプリケーションについて説明され、練習している様子を実験実施者が実演して参加者に見せた。次に、5 分間の練習直後にテスト（PreTest）を実施することを参加者に伝えた上で、アプリケーションを通してダンサーの見本映像を見ながら、ダンス動作の練習を開始してもらった。練習終了後、PreTest を実施した。PreTest では、ダンサーの見本映像（参照動画）を見ながらダンス動作を 3 回、間隔を空けて実施し、その様子を撮影した。すべてのダンス動作の練習とその PreTest を終えたら、DAY1 は終了とした（所要時間約 20 分）。実験終了後、練習中の映像から練習開始 1 分後から 3 分間分をトリミングし、学習モデルに入力し学習を開始させ、約 20 時間空けて参照動画のダンスを踊る自己映像が生成された。

DAY2 では、DAY1 と同様にダンス動作の練習とテスト（PostTest）を実施する。ただし、ダンス動作の練習時間は 10 分間とし、自己映像提示グループには、学習モデルより生成された参加者の自己映像を見ながらダンス動作を練習してもらった。他者映像提示グループでは、DAY1 と同様にダンサーの見本映像（参照動画）を見ながらダンス動作を練習してもらった。練習終了後の PostTest では、どちらのグループもダンサーの見本映像を見ながらダンス動作を 3 回、間隔を空けて実施し、その様子を撮影した。すべてのダンス動作の練習とその PostTest を終えたら、アンケートに回答してもらった。アンケートでは、それぞれのダンス動作の習得難易度を知るために、それぞれのダンス動作の習得難易度について、1: 非常に簡単, 2: かなり簡単, 3: やや簡単, 4: どちらともいえない, 5: やや難しい, 6: かなり難しい, 7: 非常に難しい、の 7 段階で回答してもらった。また参加者の自己効力感を知るために、それぞれのダンス動作に対する「今後練習を続ければダンスを習得できると思う」の質問について、全く同意できないを 1、非常に同意できるを 5 とした 5 段階のリッカート尺度で回答してもらった。参加者の達成感を知るために、それぞれのダンス動作に対する「ダンスを習得できた」の質問について、全く同意できないを 1、非常に同意できるを 5 とした 5 段階のリッカート尺度で回答してもらった。その他実験に関するコメントについては、自由記述の欄を設け、そこに回答してもらった。さらに、自己映像提示グループに対してのみ、それぞれのダンス動作ごとで自己映像を見た際に「自

## 鏡映像をリアルタイム表示

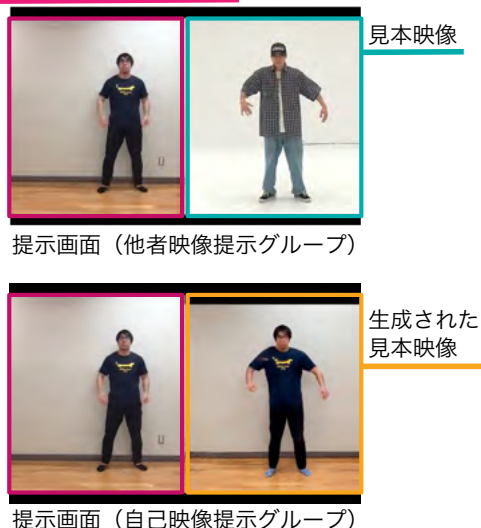


図 2 (左) グループ毎の提示画面, (右) ダンス動作例.

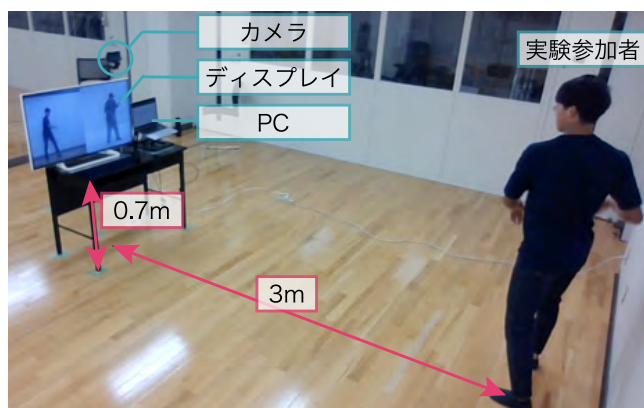


図 3 実験環境

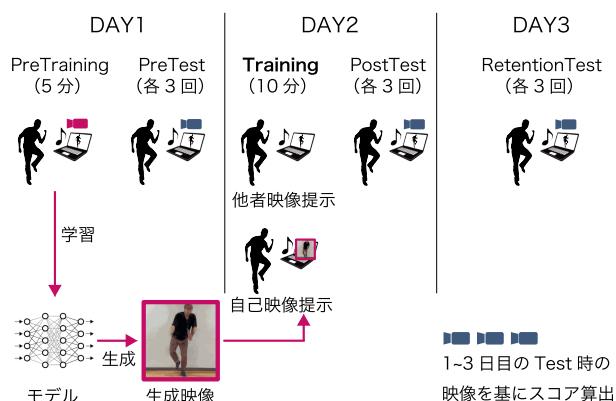


図 4 実験行程

分がダンスをしているように感じた」の質問について全く同意できないを1, 非常に同意できるを5とした5段階のリッカート尺度で回答してもらった。また自己映像を見た際の感覚などに関するコメントについては, 自由記述式のアンケートに回答してもらった。回答は Google Form を介して行ってもらった。アンケート回答後, DAY2 は終了

とした(所要時間約 45 分)。

DAY3 では, 練習の効果を保持できているか確認するために, DAY2 実施から 20 時間以上期間を空けてテスト(RetentionTest)を実施した。すべて参加者がダンサーの見本映像を見ながらダンス動作を3回, 間隔を空けて実施し, その様子を撮影した。すべてのダンス動作の RetentionTest を終わったら, DAY3 は終了とした(所要時間約 5 分)。

## 4.4 評価方法

参照動画と三つのテスト時に撮影された参加者が踊っている動画間において, 動作にどの程度差があるかのコストを次のように算出した。まず, テスト時に撮影した動画に含まれている音声と使用した楽曲との相互相関関数を求め, 最も値が高くなる箇所を開始時間として楽曲の長さ分トリミングすることで, すべてのダンス動作の開始時刻を揃えた。次に, 参照動画および切り出した動画にたいして OpenPose [14] を適用し, 図 5 に示すような 25 の特徴点(2次元の  $x, y$  座標)を算出する。

参加者の体格差を考慮し, 図 5 の赤い矢印で表現された 18 本の骨格をベクトルに見立て, それぞれ単位ベクトルに変換した。それぞれ特徴点に対して信頼度(0 ~ 1)が算出されており, 連結した特徴点の信頼度がそれぞれ 0.1 以上となるベクトルを抽出し, それ以下の場合は 0 を代入した。1 フレームあたり 36 次元のベクトルを算出し, これをすべてのフレームにおいて算出する。これらを動画一つあたりの特徴量ベクトルとし, 動画間の特徴ベクトルに対して FastDTW ライブラリ [16] を用いて DTW を計算し, 動画間の距離を算出する。これを DTW コストとし, どれだけ参照動画のダンス通りに踊れたかの評価に利用した。尚, ダンス動作は楽曲のリズムに合わせてタイミングよく



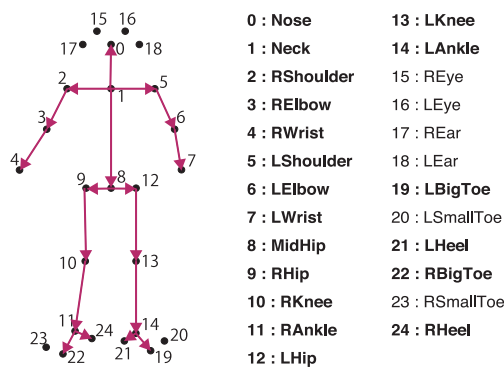


図 5 OpenPose で抽出された骨格情報から計算に利用したベクトル

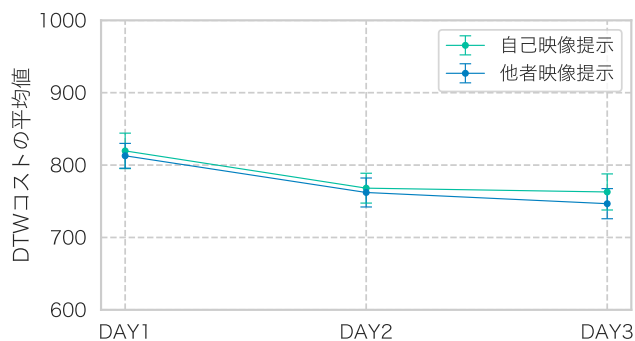


図 6 Dance1~3 の DTW コストの平均値

動くことが求められるが、今回は初学者を対象としているため、まずはどれだけ参照動画と同じ動作を行えたかを調査するために、DTW コストを評価指標としている。DTW コストが低いほど、参照動画のダンス動作と参加者のダンス動作の類似度が高いことを意味する。また、OpenPose は 2 次元座標上での骨格推定となるため、奥行き方向の動きは無視されてしまう。しかし、本実験で利用する映像生成モデルの制約から、奥行き方向の動作はあまり含まれないダンス動作を採用しているため、評価指標として問題ないとする。また今回、繰り返し再生される参照動画に合わせてテストを実施するために、ダンス動作の正確な開始タイミングから明らかに遅れて動き出す参加者が多く見られた。この影響を取り除くために、DTW コストを計算する際は最初の約 1 秒に相当する 60 フレームは飛ばして 61 フレーム目以降の特徴ベクトルを用いた。

## 5. 結果

Dance1~Dance3 における DTW コストの平均値を図 6 に示す。エラーバーは標準誤差を示す。DTW コストについて、日程 (DAY1~3) と提示グループを要因とした 2 要因の参加者間分散分析を行った結果、日程において有意な差が見られ ( $F(2, 174) = 5.75, p < 0.01$ )、提示グループには有意差が見られなかった。Holm 法による多重比較を行った結果、DAY1 と DAY2、DAY1 と DAY3 の間にそれぞれ

有意差が見られた ( $p < 0.05$ )。次に、ダンスの種類ごとの DTW コストの平均値を図 7 に示す。エラーバーは標準誤差を示す。Dance1 の DTW コストについて、日程と提示グループを要因とした 2 要因の参加者間分散分析を行った結果、日程において有意が見られ ( $F(2, 54) = 4.07, p < 0.05$ )、提示グループには有意差は見られなかった。Holm 法による多重比較を行った結果、DAY1 と DAY3 の間に有意差が見られた ( $p < 0.05$ ) 次に Dance2 の DTW コストについて、日程と提示グループを要因とした 2 要因の参加者間分散分析を行った結果、日程において有意が見られ ( $F(2, 54) = 5.80, p < 0.01$ )、提示グループには有意差は見られなかった。Holm 法による多重比較を行った結果、DAY1 と DAY3 の間に有意差が見られた ( $p < 0.05$ ) また Dance3 の DTW コストについて、日程と提示グループを要因とした 2 要因の参加者間分散分析を行った結果、2 つの要因においてそれぞれ有意差は見られなかった。

次に、アンケート結果について確認する。図 8 は、ダンス動作の習得難易度に関する回答結果を示しており、エラーバーは標準誤差を示す。以降の図のエラーバーについても同じく標準誤差を示している。2 要因参加者間計画で提示グループと Dance の種類をそれぞれ要因とした分散分析を行った結果、Dance の種類の間で有意差が見られた ( $F(2, 54) = 73.50, p < .01$ )。Holm 法を用いて多重比較を行ったところ、ダンス間のすべての組み合わせにおいて有意差を確認した ( $p < .05$ )。よって、Dance1、Dance3、Dance2 の順に段階的に難しいダンス動作になっていることがわかる。一方、提示グループの間には有意な差は見られなかった。図 9 は、「今後練習を続ければダンス動作を習得できると思う」に関する回答結果を示しており、参加者の自己効力感を示すと考える。2 要因参加者間計画で提示グループと Dance の種類をそれぞれ要因とした分散分析を行った結果、提示グループで有意傾向があり ( $F(1, 54) = 3.904, p < 0.1$ )、Dance の種類で有意差が見られた ( $F(2, 54) = 61.51, p < 0.01$ )。提示グループについて、自己映像提示グループの平均 2.8 が他者映像提示グループの平均 3.2 よりも有意に小さい傾向があった。Holm 法を用いて多重比較を行ったところ、ダンス間のすべての組み合わせにおいて有意差を確認した ( $p < 0.05$ )。よって、Dance1、Dance3、Dance2 の順にダンス動作を習得しやすいと参加者は感じていることがわかる。一方、Dance3 については平均 0.5 以上の差がみられるものの、提示グループの間には有意な差は見られなかった。図 10 は、「ダンスを習得できた」に関する回答結果を示しており、参加者の達成感を示す。2 要因参加者間計画で提示グループと Dance の種類をそれぞれ要因とした分散分析を行った結果、Dance の種類の間で有意差が見られた ( $F(2, 54) = 35.65, p < .01$ )。Holm 法を用いて多重比較を行ったところ、ダンス間のすべての組み合わせにおいて有意差を確認した ( $p < .05$ )。

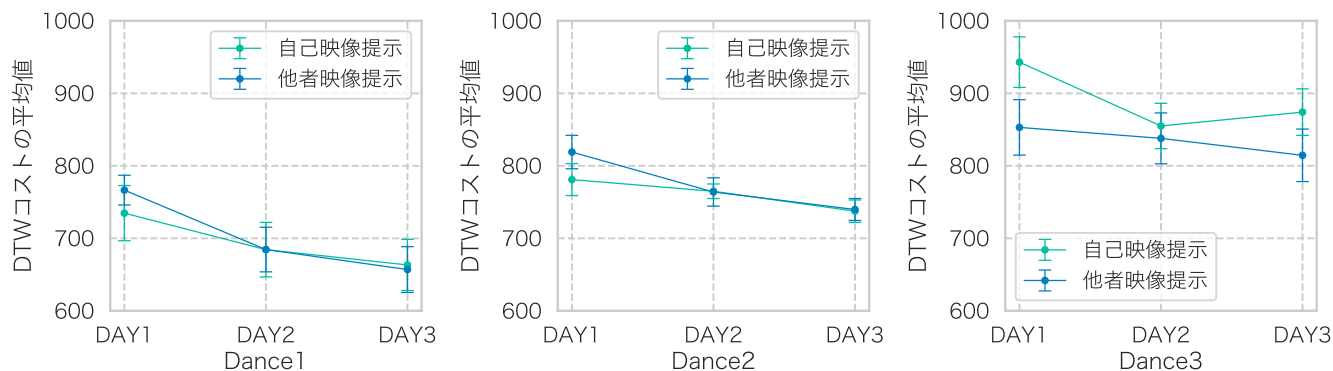


図 7 Dance ごとの DTW コストの平均値

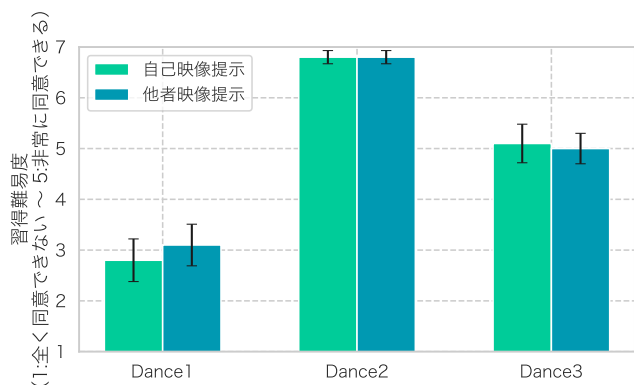


図 8 「ダンスの習得難易度について」に対する回答結果

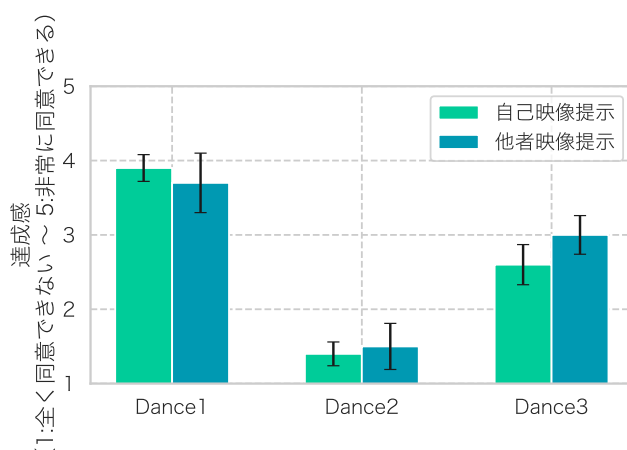


図 10 「ダンスを習得できた」に対する回答結果

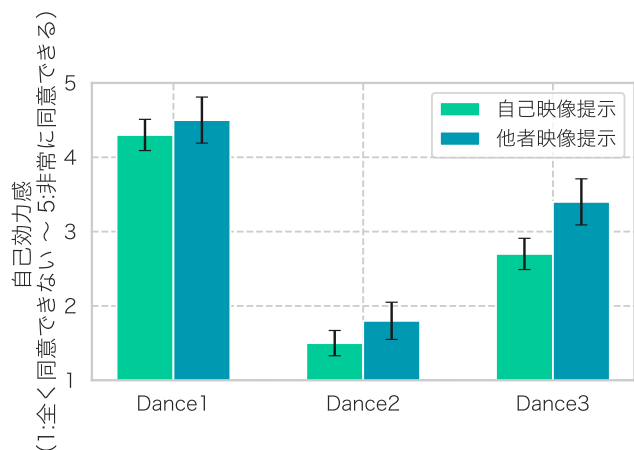


図 9 「今後練習を続ければダンスを習得できると思う」に対する回答結果

よって、Dance1 は他と比べて高い達成感を得られており、逆に Dance2 は低いことがわかる。一方、提示グループの間には有意な差は見られなかった。図 11 は、自己映像提示グループに対してのみ実施したアンケート結果である。「自分がダンスをしているように感じた」の文について全く同意できないを 1、非常に同意できるを 5 とする 5 段階での評価を行ってもらった。図 8 の習得難易度の結果を考慮すると、簡単なダンス動作ほど自分が踊っているように

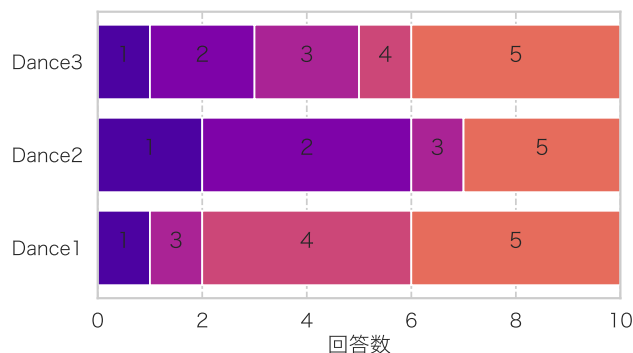


図 11 「自分がダンスをしているように感じた」に対する回答結果

感じる傾向が見られた。

次に自由記述について確認する。ポジティブなコメントは多数見られた、特に動きの把握や動かし方について、「踊っている自分の姿を実際に見ることができるので、どのように動いたらいいかを掴みやすかった」、「もとの映像はダンスしている人がどのような動きをしているのか理解できなかった。自己映像では複雑な動きが少し簡単になっているように感じた」、「他人がやっているよりも、体の動かし方が分かりやすく感じました」、PreTraining 時「の練習より自分の動きとお手本との違いに気づけた気がする」、

「体のキレや動かすタイミング、などが分かりやすいと思った」、「体の動かし方が理解しやすくなったように感じた」、「今の自分の動きと比較しやすい」、「自己映像と現在の自分の姿が横並びに映し出されているので、お手本の映像よりも動きの比較と修正が容易かった」、「自分がどのように動けば良いのか、自分のダンスとどう違うのか分かりやすかった」、「体型が同じなので体の各部位の動きを比べやすい気がした」、などのコメントが見られ、自己映像の方が動きの理解を向上させる可能性がある。

また、「自分が本来できないはずの動きをしている映像だったので不思議な感じはあったが、動きのイメージがしやすかった」、「今の自分も映像のように踊れていると思いながら体を動かしていた。たまに今の自分の鏡映像を見て比較したらうまく動けていないポイントに気づけた。違和感はないかった」、「自分が上手く動けたらこんなふうに見えるのか、というイメージがつきやすい」などのコメントもあり、動作のイメージをサポートできる可能性が示された。

さらに、「自分がこんなに踊れている映像を見たことがなかったので単純におもしろかった」、「自分が完璧に踊っている映像が違和感バリバリで面白かった」、「自分がうまくなった様子が分かって、モチベーションは上がりました」、「上手く踊れている自分が映像で見えるので、上手く踊っていると錯覚して楽しく練習できる」などのコメントがあり、面白さという点で学習者のモチベーションを高められる可能性がある。

他に、「見本が同じ人物・同じ場所だったため、どう真似をすればいいかのイメージが立てやすかったです」とのコメントがあり、今回は学習する場所と同じ場所で撮影された映像を用いたが、背景が変化すると異なる効果を与える可能性がある。

ネガティブなコメントとしては、「お手本の映像と比べてノイズが入っているところもあったので、細かい動きがわからないところもあった」、「下半身・足の動きが分かりにくかったです」、「ダンス1の映像で、足がどちらが前に出ているかがかなり分かりにくくなっていました。そもそも難しい動きがある場合は、自己映像があっても、イメージはしにくい印象でした。合成によって着ている服の色によっては奥行きが見にくいように感じた」、「画質があまり良くなかったため、手先などの細かい部分を確認することが難しかったです」などの映像生成の質の低さに関する指摘があり、これらは改善の必要がある。

また、今回は実験条件を統制するために、テスト時はこちらのグループも参照動画を見ながら実施してもらったが、「実際の映像と微妙に動きが違う気がしたテストの時、その差によって少し頭がこんがらがったかもしれない」、「テストの時の映像が違うダンサーに戻って、少し焦る。」などの指摘があり、テストを実施する前に、参照動画に慣れて

もらう時間を設けた方が、参加者の混乱を避けられた可能性がある。

## 6. 考察

図6のDance1~3のDTWコストの平均値の結果から、参加者の学習を進めることに貢献できているが、提示グループ間に有意な差はなかった。考えられる原因の一つとして、自由記述で見られるように生成した映像の質の低さが、ダンス動作の学習に影響している可能性がある。映像の不自然さを取り除いた、より質の高い映像の生成が求められることがわかった。Danceの種類それぞれについても、提示グループの間には有意な差は見られなかった。難易度による影響はないか、引き続き調査していく必要がある。また、表示する画面のサイズが小さかった(90 cm × 50 cm)ために、鏡としての役割が薄まり、セルフモデリングの効果を得られなかった可能性がある。Elsayedら[17]が示すように、比較的大きなディスプレイを使用した方が姿勢誘導の際のエラー率を低減させられることがわかっている。より大きなディスプレイを用いることで、今回のアンケートの自由記述コメントで得られた自己映像に含まれるメリットが定量的に測定できる効果として現れる可能性はある。

図9の結果から、自己映像提示グループは他者映像提示グループより自己効力感が有意に小さい傾向があり、自己映像の提示は自己効力感を低下させる可能性がある。自由記述では、「自分が映像ぐらい踊れていると錯覚してしまう。そのため鏡映像とこまめに比較しないと動きが違ってくるような気がした」、「実際には踊れてないのに自分が踊っていたので変な感じでした」などの意見が見られ、自分がそもそもできない動作、現段階ですぐには到達できないような動作を自己映像で提示したために、評価が低下したのではないかと考える。一方で「少し違和感がある部分はあったが、自分の顔で上手なダンスをしているようにちゃんと感じれて、ちゃんとこのダンス上手く踊れるかもと感じれた」、「他人がやっているバージョンだと全く真似できなさそうだったんですが、自己映像で自分がダンスしてる姿を見ると、再現できる可能性を感じました」、「練習していけば、自己映像のようになるイメージはしやすいと感じました」といったコメントもあり、人によっては自己効力感を高める可能性もあり、今後も引き続き調査が必要と考えている。

図11に示す結果では、回答についてバラつきが見られ、参加者によっては生成された映像を見ても自分が踊っているようには全く感じないと回答している。自由記述では、「身体が小さくなって」いる、「足先が少し見えないことがあった」、「首がなくなってしまったり、顔の向きがおかしかったりといった点で不自然さを感じました」、「体格が合っていないように感じます」、「後半が画質が悪くなってしまっ」あまり踊っているように感じないといった指摘が

あった。ビデオセルフモデリングでは、自身が行なっているように錯覚させることが前提となっており、ビデオセルフモデリングの効果を得るためには、自身が動作を実施している感覚を強めることが望ましい。そのため、先ほど列挙したような不自然さは排除していく必要がある。また、図 8 と図 11 に示す結果から、学習する動作に対して難しさを感じるほど、自分が踊っているような感覚が弱くなることがわかる。この点については、学習者のスキルレベルに合わせて対戦相手 [13] や、利用する道具 [18] を調整していくような手法と同様に、学習者のダンス動作レベルに合わせた自己映像を提示することが望ましい。例えば、ダンスのエキスパートの動きをマスターしている自己映像を最初に提示するのではなく、エキスパートに一步近づいたような自己映像を、学習者のレベルに合わせて段階的に提示できる技術、エキスパートのスキルと初学者のスキルをモーフィングできるような技術の開発が今後必要になると考えている。

## 7. まとめ

本研究では、ダンス動作をマスターした自身の映像を先に見ることによるダンス学習支援手法を提案した。練習時の映像を入力とし、深層学習を用いて参照動画中のダンサーの動作と同じ動作を行う自己映像を生成し、提示する。自動生成した自己映像が、ダンス学習に有効かどうか検証するために、参加者 20 名の評価実験を実施した。実験の結果、単に参照動画を見る場合と比べて有意な差は見られなかった。また、アンケート結果から、自己映像を提示されたグループの自己効力感が有意に低い傾向が見られた。

**謝辞** 本研究の一部は、JST CREST(JPMJCR18A3)の支援によるものである。ここに記して謝意を表す。

## 参考文献

- [1] Ward, P., Hodges, N. J. and Williams, A. M.: Deliberate Practice and Expert Performance: Defining The Path to Excellence, *Journal of Skill acquisition in sport*, pp. 255–282 (2004).
- [2] Peeters, T., van breda, E., Saeys, W., Schaerlaken, E., Vleugels, J., Truijen, S. and Verwulgen, S.: Vibrotactile Feedback During Physical Exercise: Perception of Vibrotactile Cues in Cycling, *International Journal of Sports Medicine*, Vol. 40 (2019).
- [3] Landry, S. and Jeon, M.: Interactive Sonification Strategies for the Motion and Emotion of Dance Performances, *Journal of Multimodal User Interfaces*, Vol. 14 (2020).
- [4] Kyan, M., Sun, G., Li, H., Zhong, L., Muneesawang, P., Dong, N., Elder, B. and Guan, L.: An Approach to Ballet Dance Training through MS Kinect and Visualization in a CAVE Virtual Reality Environment, *Journal of ACM Transactions on Intelligent Systems and Technology*, Vol. 6, No. 2 (2015).
- [5] Choi, J.-H., Lee, J.-J. and Nasridinov, A.: Dance Self-Learning Application and Its Dance Pose Evaluations, *Proceedings of the 36th Annual ACM Symposium on Applied Computing, SAC '21*, pp. 1037–1045 (2021).
- [6] Fujimoto, M., Terada, T. and Tsukamoto, M.: A Dance Training System that Maps Self-Images onto an Instruction Video, *Proceedings of the 5th International Conference on Advances in Computer-Human Interactions, ACHI '12*, pp. 309–314 (2012).
- [7] Creer, T. L. and Miklich, D. R.: The Application of a Self-modeling procedure to Modify Inappropriate Behavior: A Preliminary Report, *Behaviour Research and Therapy*, Vol. 8, No. 1, pp. 91–92 (1970).
- [8] Ste-Marie, D., Vertes, K., Rymal, A. and Martini, R.: Feedforward Self-Modeling Enhances Skill Acquisition in Children Learning Trampoline Skills, *Journal of Frontiers in Psychology*, Vol. 2, p. 155 (2011).
- [9] Chan, C., Ginosar, S., Zhou, T. and Efros, A. A.: Everybody Dance Now, *Proceedings of the IEEE/CVF International Conference on Computer Vision, ICCV '19* (2019).
- [10] Clark, S. E. and Ste-Marie, D. M.: The Impact of Self-as-a-model Interventions on Children's Self-regulation of Learning and Swimming Performance, *Journal of Sports Sciences*, Vol. 25, No. 5, pp. 577–586 (2007).
- [11] Dowrick, P.: Self Model Theory: Learning From the Future, *Wiley Interdisciplinary Reviews: Cognitive Science*, Vol. 3, pp. 215–230 (2012).
- [12] Steel, K., Mudie, K., Sandoval, R., Anderson, D., Dogramaci, S., Rehmanjan, M. and Birznieks, I.: Can Video Self-Modeling Improve Affected Limb Reach and Grasp Ability in Stroke Patients?, *Journal of motor behavior*, Vol. 50, pp. 1–10 (2017).
- [13] Michael, A. and Lutteroth, C.: Race Yourself: A Longitudinal Exploration of Self-Competition Between Past, Present, and Future Performances in a VR Exergame, *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20*, pp. 1–17 (2020).
- [14] Cao, Z., Hidalgo Martinez, G., Simon, T., Wei, S. and Sheikh, Y. A.: OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2019).
- [15] Tsuchida, S., Fukayama, S., Hamasaki, M. and Goto, M.: AIST Dance Video Database: Multi-genre, Multi-dancer, and Multi-camera Database for Dance Information Processing, *Proceedings of the 20th International Society for Music Information Retrieval Conference, ISMIR '19*, pp. 501–510 (2019).
- [16] Salvador, S. and Chan, P.: Toward Accurate Dynamic Time Warping in Linear Time and Space, *Journal of Intelligent Data Analysis*, Vol. 11, No. 5, pp. 561–580 (2007).
- [17] Elsayed, H., Hoffmann, P., Günther, S., Schmitz, M., Weigel, M., Mühlhäuser, M. and Müller, F.: CameraReady: Assessing the Influence of Display Types and Visualizations on Posture Guidance, *Designing Interactive Systems Conference 2021, DIS '21*, pp. 1046–1055 (2021).
- [18] Turakhia, D. G., Wong, A., Qi, Y., Blumberg, L.-G., Kim, Y. and Mueller, S.: Adapt2Learn: A Toolkit for Configuring the Learning Algorithm for Adaptive Physical Tools for Motor-Skill Learning, *Designing Interactive Systems Conference 2021, DIS '21*, pp. 1301–1312 (2021).