

第67回産総研人工知能セミナー「HCI × AI : 人工知能で進化するHCI研究の最前線」
2023/07/25

Deepfake Video Self-modeling

お茶の水女子大学 講師
土田修平

Shuheï Tsuchida, Haomin Mao, Hideaki Okamoto, Yuma Suzuki, Rintaro Kanada, Takayuki Hori, Tsutomu Terada, and Masahiko Tsukamoto. 2022. Dance Practice System that Shows What You Would Look Like if You Could Master the Dance. In Proceedings of the 8th International Conference on Movement and Computing (MOCO '22), Article 15, 1–8. <https://doi.org/10.1145/3537972.3537991>

自己紹介

名前: 土田 修平

経歴:

2017年 3月 神戸大学大学院工学研究科博士課程後期課程
電気電子工学専攻 修了

2017年 4月 国立研究開発法人 産業技術総合研究所 特別研究員

2019年12月 神戸大学大学院工学研究科 特命助教

2022年11月 神戸大学 未来医工学研究開発センター 特命講師

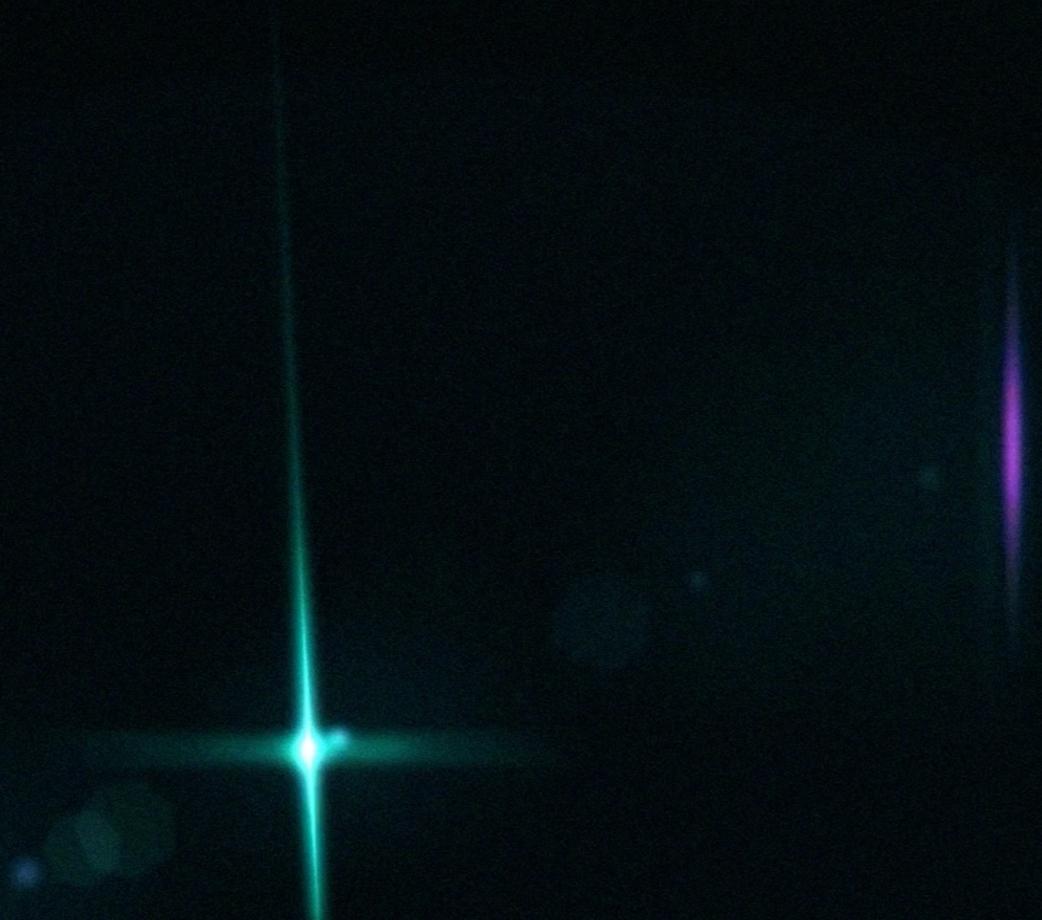
2023年 4月 お茶の水女子大学 文理融合・AIデータサイエンスセンター 講師
(来年度より共創工学部に異動)

専門分野 :

ヒューマンコンピュータインタラクション, ダンス情報処理
ウェアラブル・ユビキタスコンピューティング

自己紹介

名前: 土田 修平



自己紹介

名前: 土田 修平

経歴:

2017年 3月 神戸大学大学院工学研究科博士課程後期課程
電気電子工学専攻 修了

2017年 4月 国立研究開発法人 産業技術総合研究所 特別研究員

2019年12月 神戸大学大学院工学研究科 特命助教

2022年11月 神戸大学 未来医工学研究開発センター 特命講師

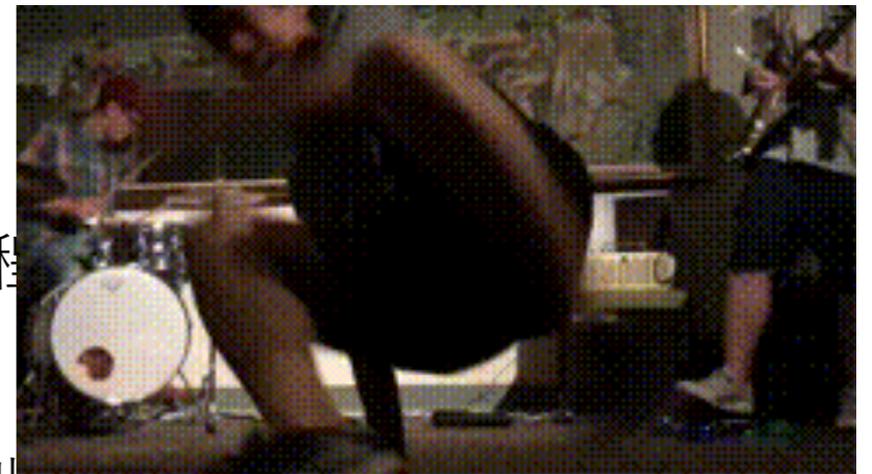
2023年 4月 お茶の水女子大学 文理融合・AIデータサイエンスセンター 講師
(来年度より共創工学部に異動)

専門分野 :

ヒューマンコンピュータインタラクション, **ダンス情報処理**
ウェアラブル・ユビキタスコンピューティング

自己紹介

名前: 土田 修平



総

大
電子

課程

2017年 4月 国立研究開発法人 産業技術総合研究所 特別研究員

2019年



究科 物

2022年

究開発

2023年

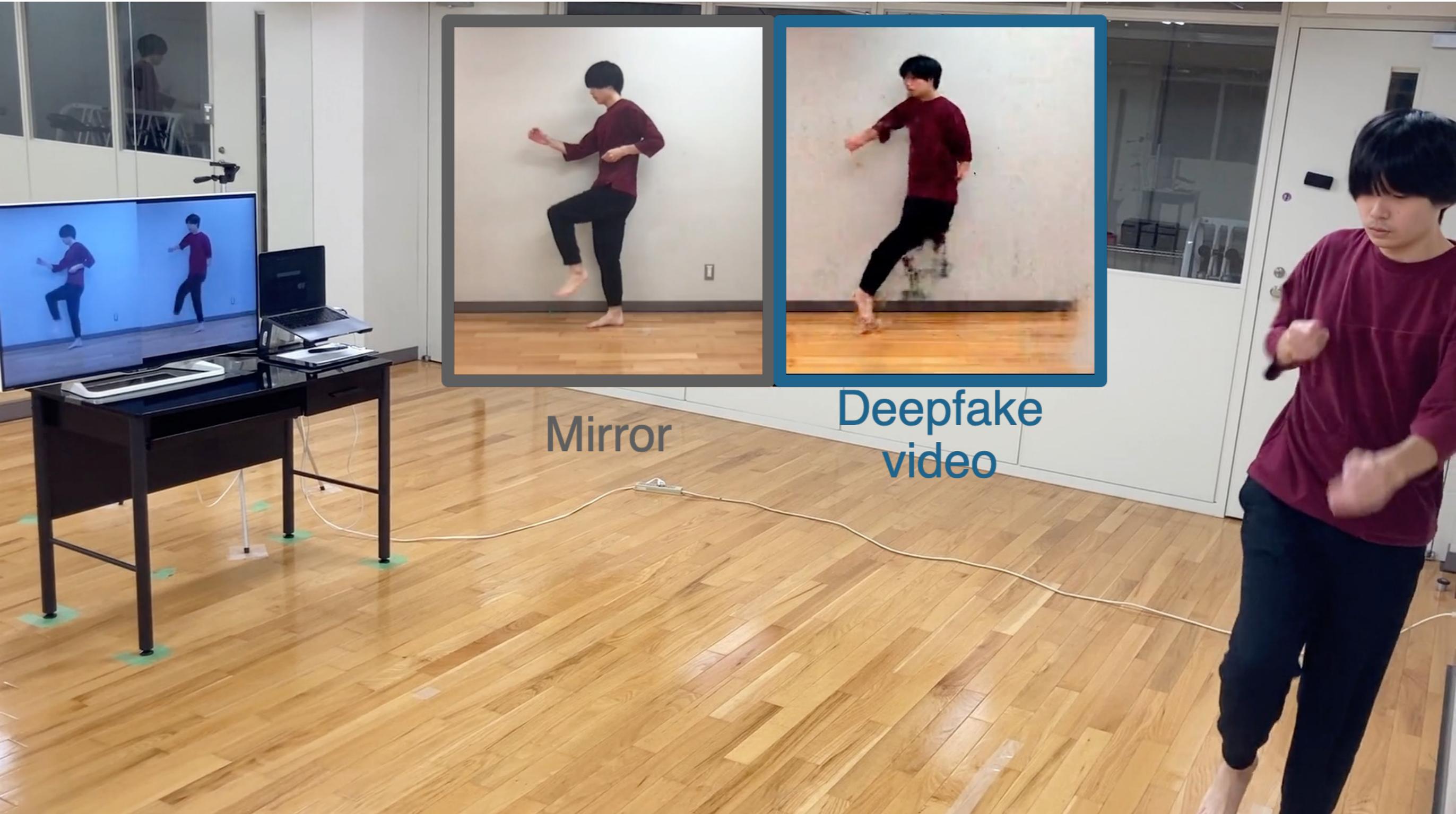
融合
部に異



専門分野 :

ヒューマンコンピュータインタラクション, **ダンス情報処理**
ウェアラブル・ユビキタスコンピューティング

研究デモ



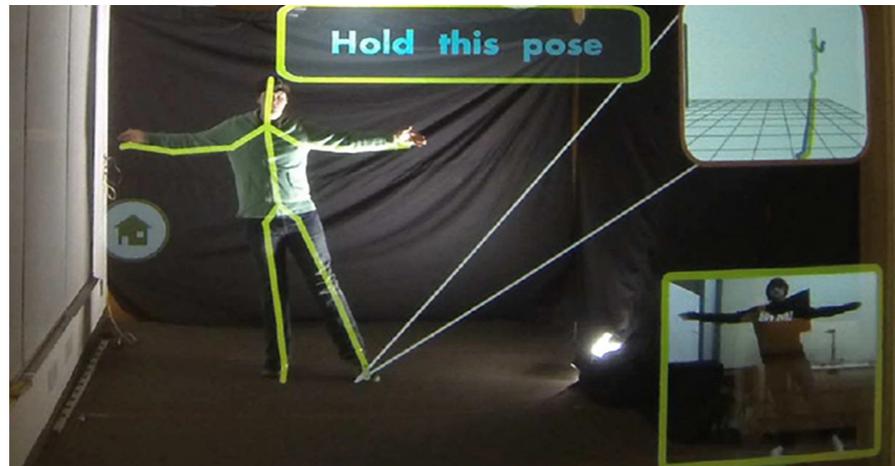
Mirror

Deepfake
video

背景

ダンス動作の習得

ダンス動作の習得を支援する研究は多数みられる



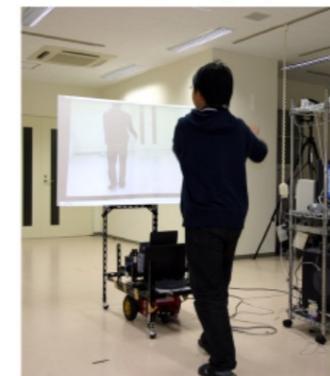
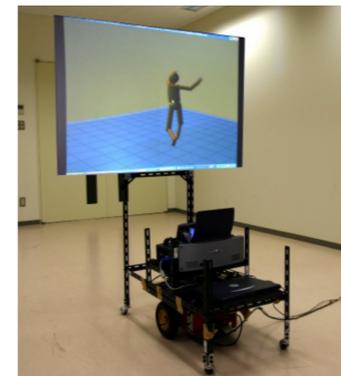
鏡の拡張
[Andresonら, 2013]



聴覚フィードバック
[Großhauserら, 2012]



触覚フィードバック
[Schönauerら, 2012]



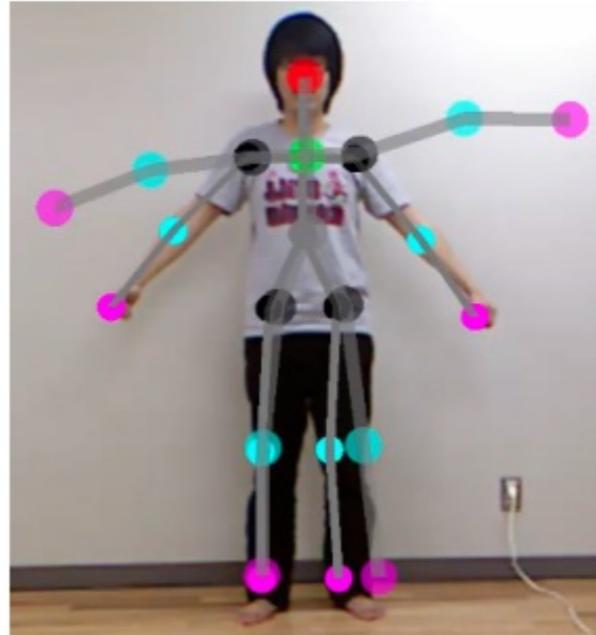
ロボット
[Nakamuraら, 2005]

ダンスを上手く踊っている自分自身を提示する練習手法 [1]

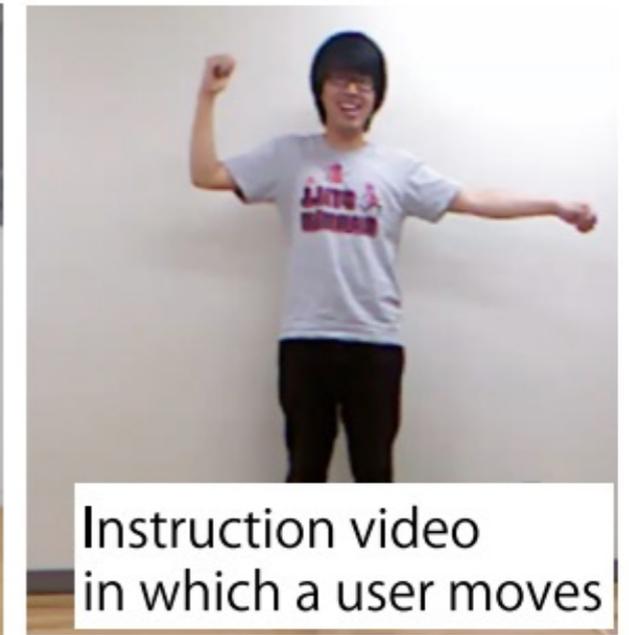
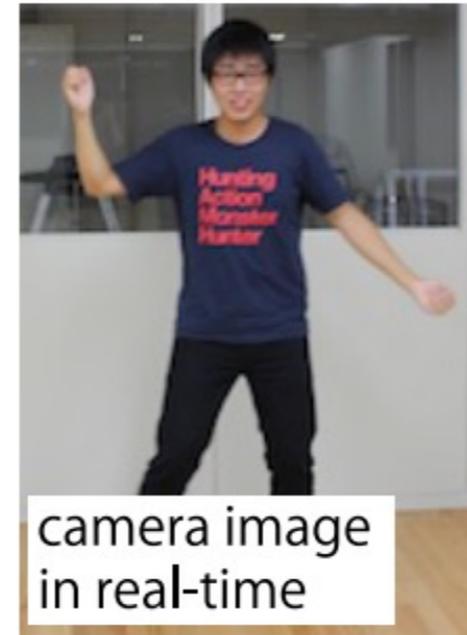
1. Create an instruction video



2. Create a self-motion-based instruction video



3. A motion training system using self-motion-based instruction video



リハビリテーション[2]やスポーツでのスキル学習[3]でも活用

→ 効果的であることが報告されている

[1] Minoru Fujimoto, Tsutomu Terada, and Masahiko Tsukamoto: A Dance Training System that Maps Self-Images onto an Instruction Video, Proceedings of the Fifth International Conference on Advances in Computer-Human Interactions (ACHI 2012), pp. 309–314 (2012).

[2] Kylie Steel, Kurt Mudie, Remi Sandoval, David Anderson, Sera Dogramaci, Mohammad Rehmanjan, and Ingvars Birznieks: Can Video Self-Modeling Improve Affected Limb Reach and Grasp Ability in Stroke Patients?, Journal of motor behavior, Vol. 50, pp. 1–10 (2017).

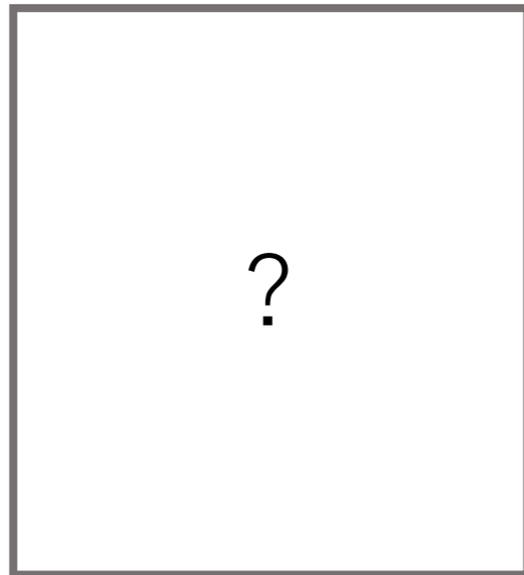
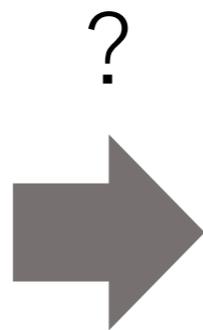
[3] Diane Ste-Marie, Kelly Vertes, Amanda Rymal, and Rose Martini.: Feedforward Self-Modeling Enhances Skill Acquisition in Children Learning Trampoline Skills. Journal of Frontiers in Psychology, Vol. 2, p. 155 (2011).

ビデオセルフモデリングの問題点

理想的な動作を行う映像の生成に手間暇がかかる



自身の練習映像

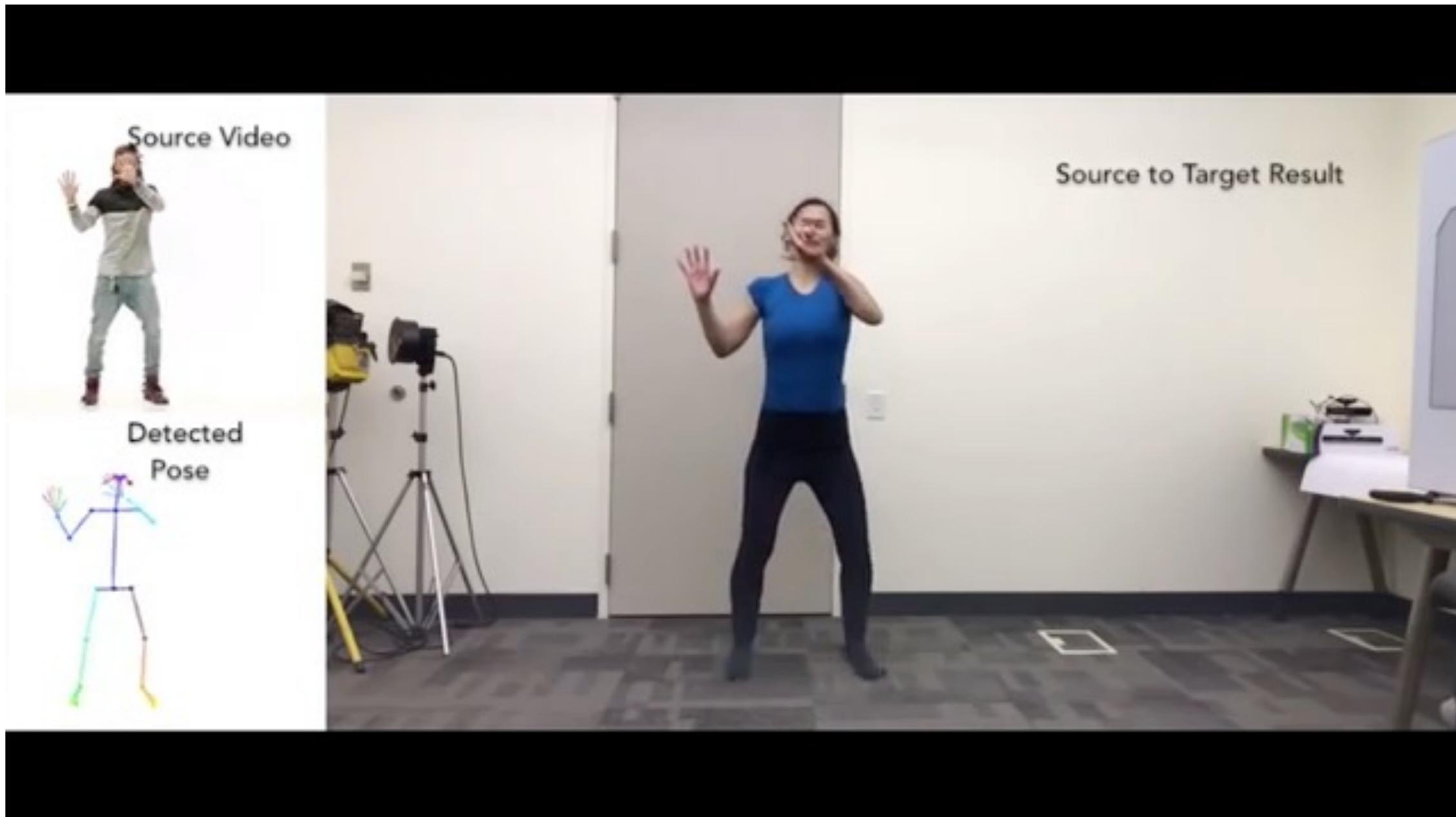


理想的な動作の映像



見本映像

Everybody dance now [1]



<https://youtu.be/PCBTZh41Ris>

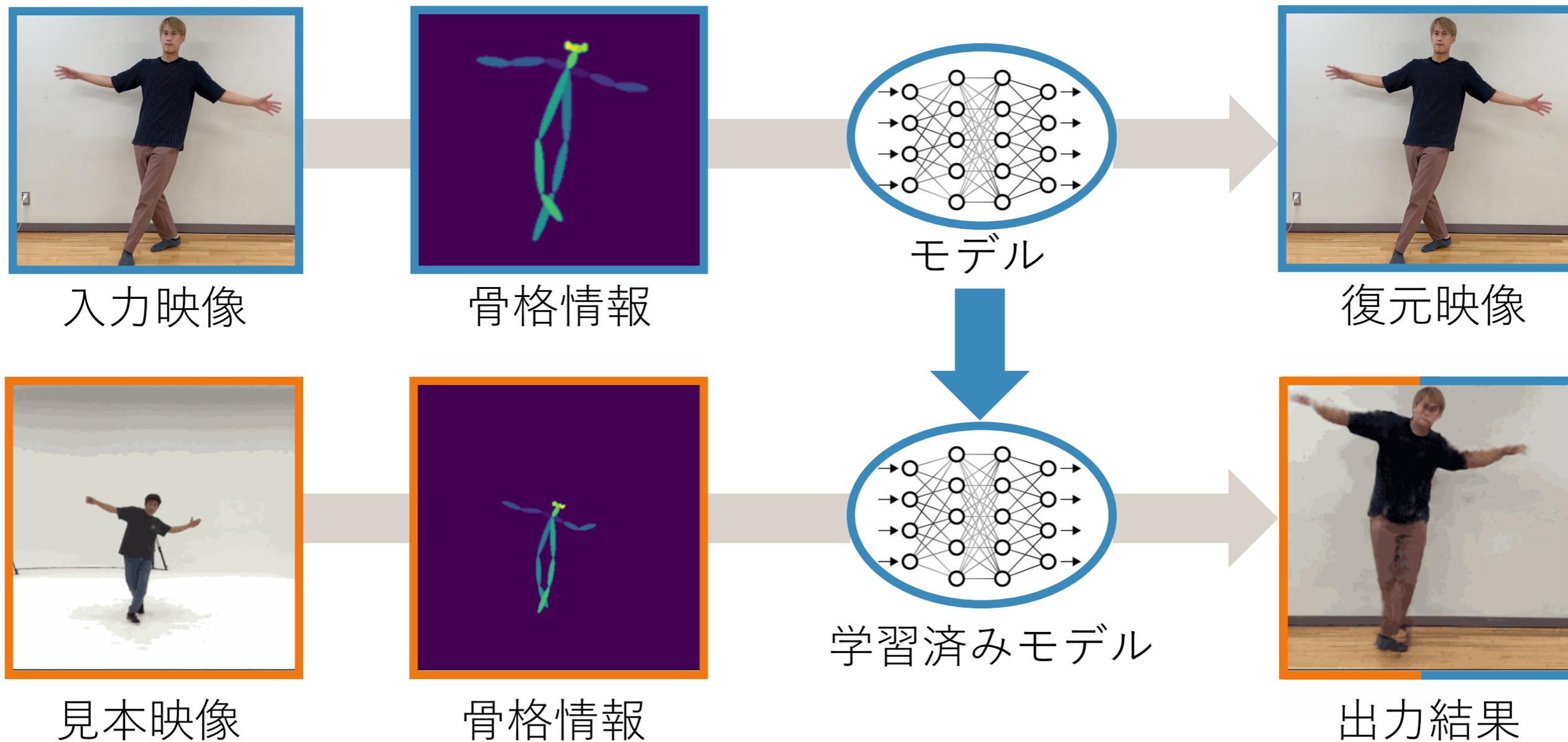
[1] Caroline Chan, Shiry Ginosar, Tinghui Zhou, Alexei A. Efros: Everybody Dance Now, Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 5933–5942 (2019).

深層学習を用いてダンスをマスターした自身の映像を生成し
その映像を先に見ることによるダンス学習支援

提案手法

見本映像生成手法

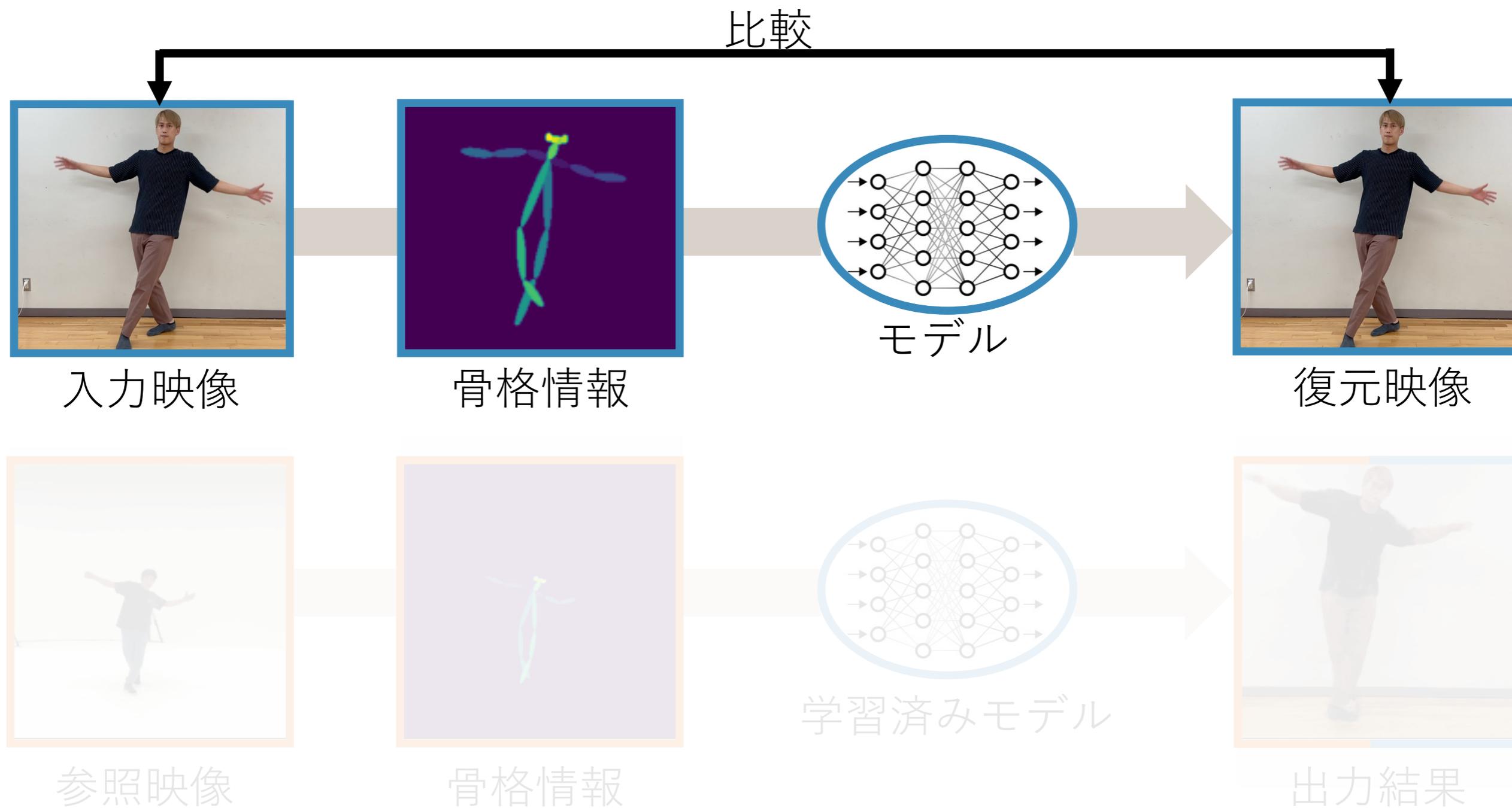
Everybody dance now[1]論文の公式ライブラリを活用



[1] Caroline Chan, Shiry Ginosar, Tinghui Zhou, Alexei A. Efros: Everybody Dance Now, Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 5933–5942 (2019).

見本映像生成手法

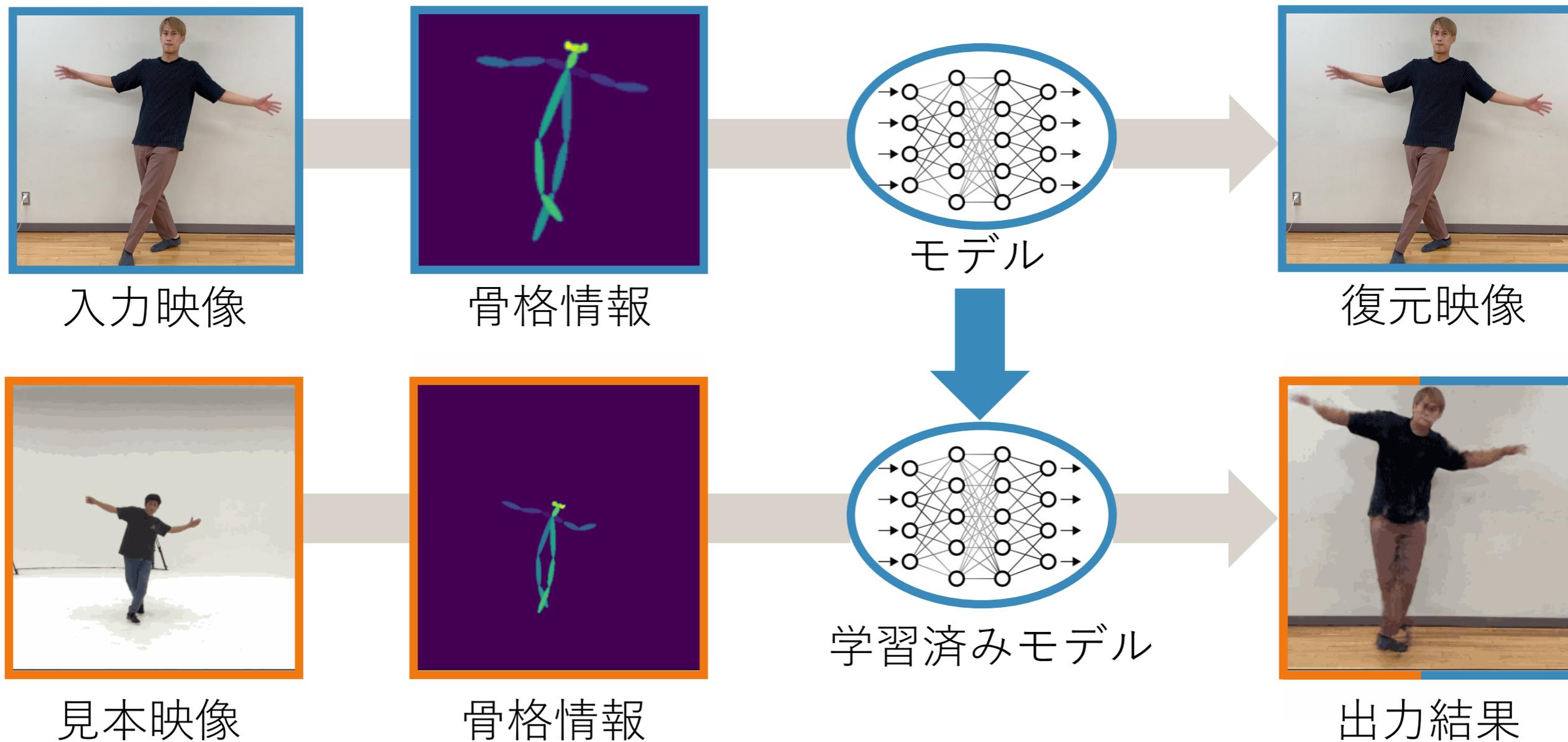
Everybody dance now[1]論文の公式ライブラリを活用



[1] Caroline Chan, Shiry Ginosar, Tinghui Zhou, Alexei A. Efros: Everybody Dance Now, Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 5933–5942 (2019).

見本映像生成手法

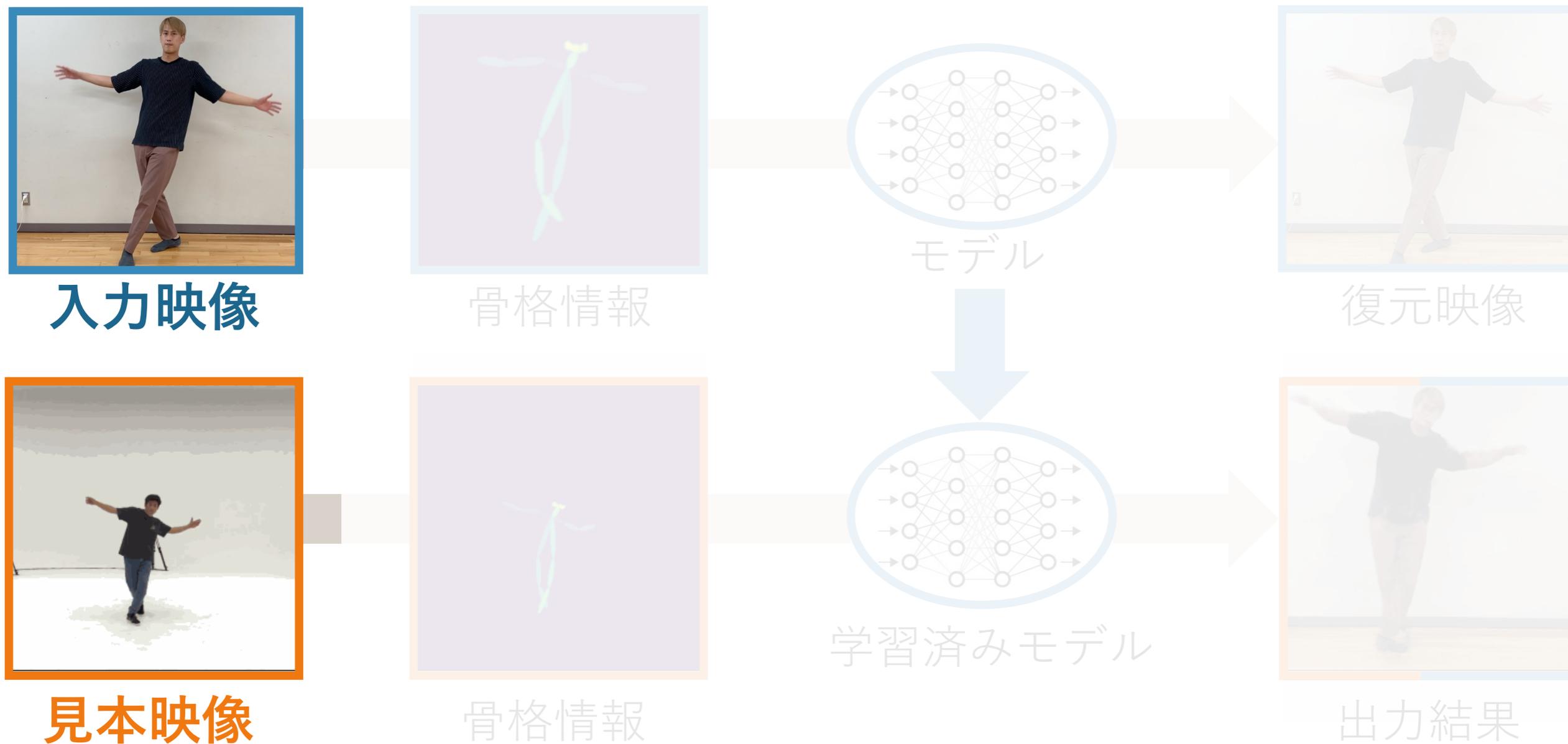
Everybody dance now[1]論文の公式ライブラリを活用



[1] Caroline Chan, Shiry Ginosar, Tinghui Zhou, Alexei A. Efros: Everybody Dance Now, Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 5933–5942 (2019).

見本映像生成手法

Everybody dance now[1]論文の公式ライブラリを活用



[1] Caroline Chan, Shiry Ginosar, Tinghui Zhou, Alexei A. Efros: Everybody Dance Now, Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 5933–5942 (2019).

予備調査 { ① 入力映像の動作の種類
② 見本映像の動作の特徴

見本映像生成の検討①: 入力映像の動作の種類 8

自由な動作



見本映像と同じ動作



入力

出力

見本映像生成の検討①: 入力映像の動作の種類 8

自由な動作



見本映像と同じ動作

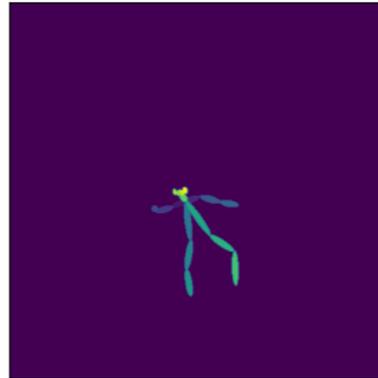


入力

出力

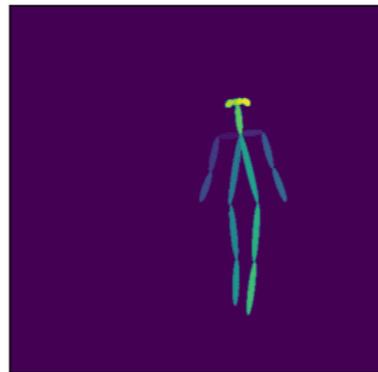
見本映像と同じ動作を行なっている映像を学習した方が出力映像のクオリティが高い

手を回す



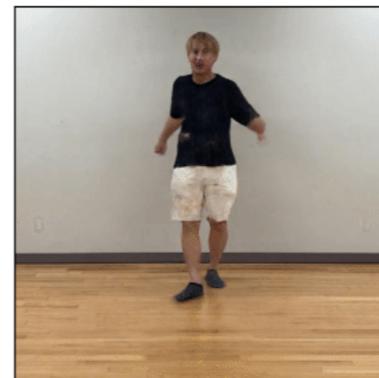
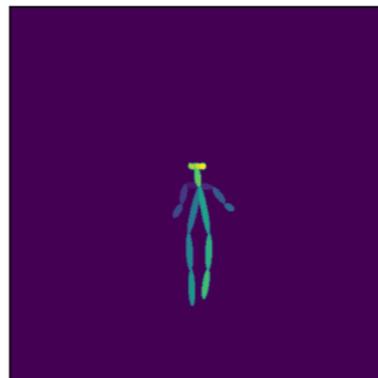
× 腕が消える

上下方向



× 立ち位置が変わる

前後方向



× 奥行き表現

これら動作を含まない振り付けで評価実験を実施

評価実験

実験目的

深層学習を用いて生成した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検証

学習対象のダンス動作

3つのダンス動作（難易度: 易 - 中 - 難）

被験者

大学生20名（男性19名，女性1名）

ダンスの経験は中学・高校での授業・イベントに参加した程度

提示グループ

他者映像提示グループ（従来手法）

自己映像提示グループ（提案手法）

- 奥行き方向の動作を含まない
- 前後方向の動作を含まない
- 腕を回す動作を含まない
- 後ろを向いていない



Dance 2

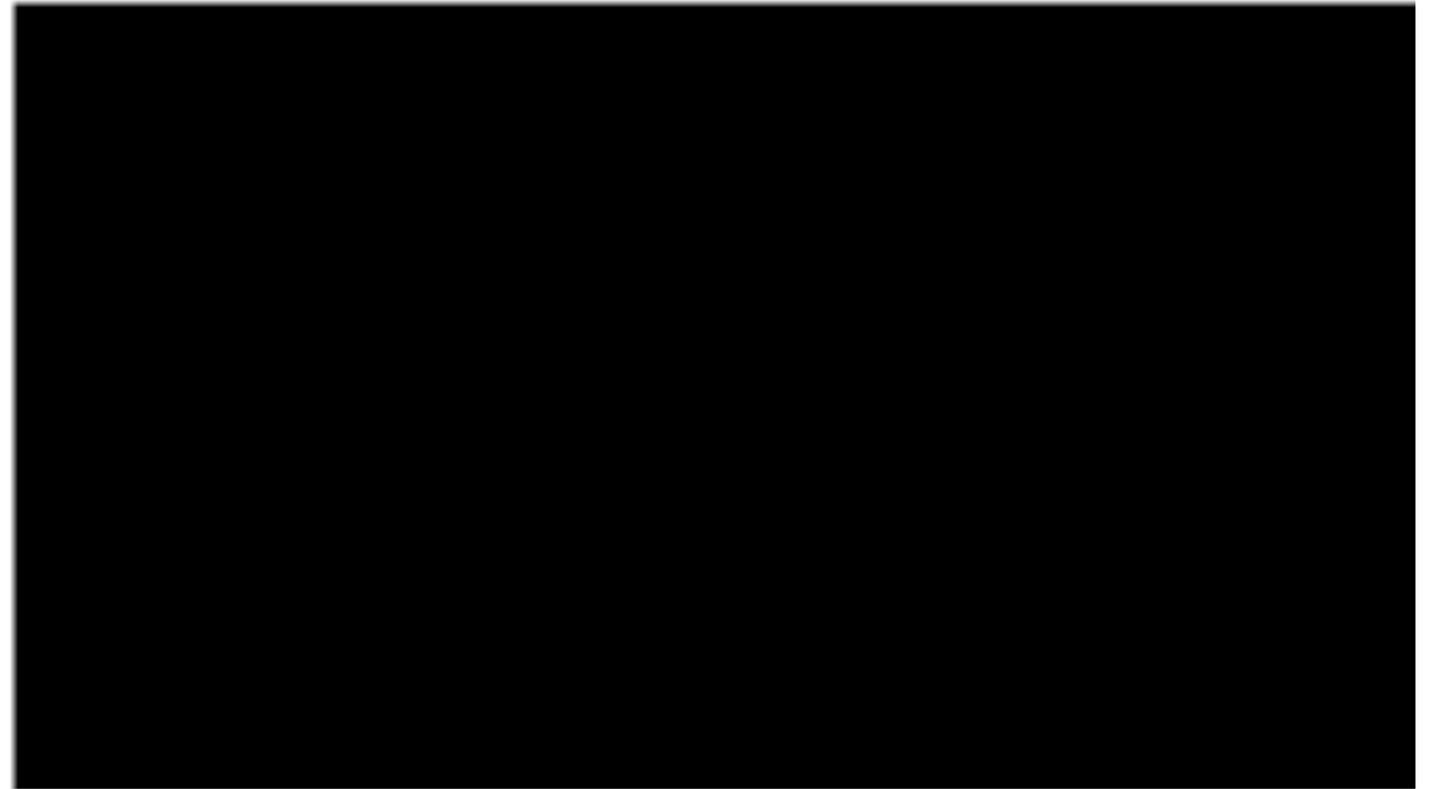


Dance 1

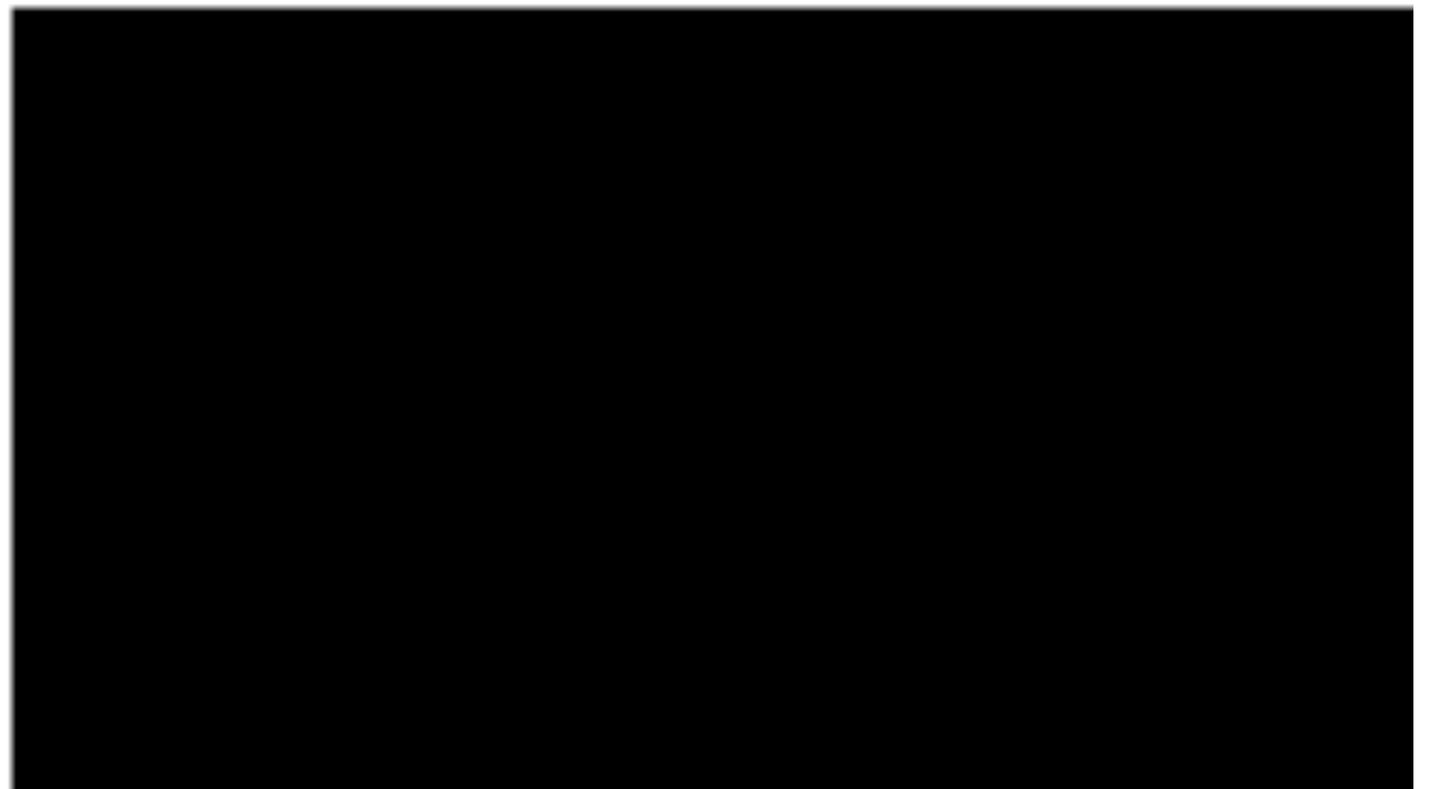


Dance 3

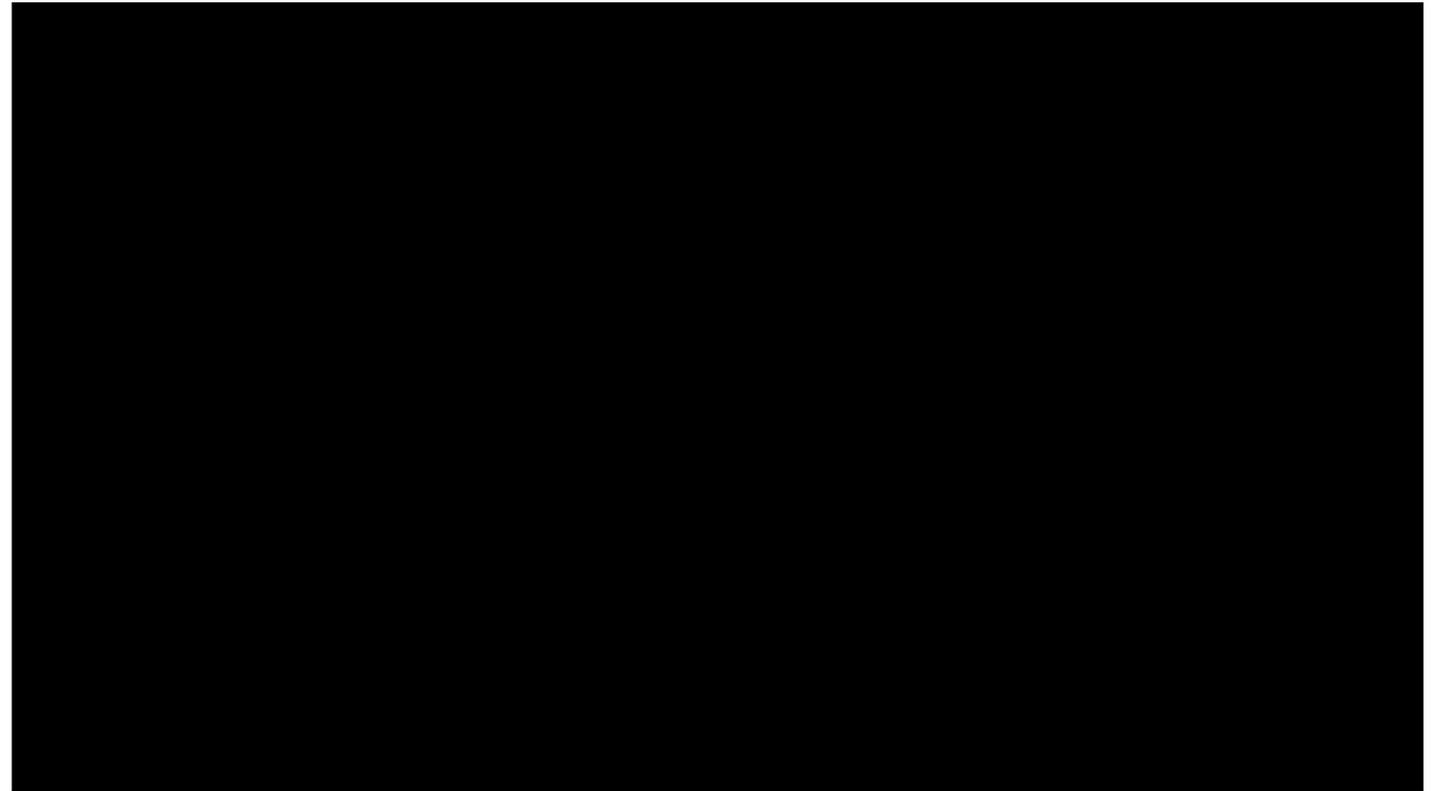
他者映像提示グループ



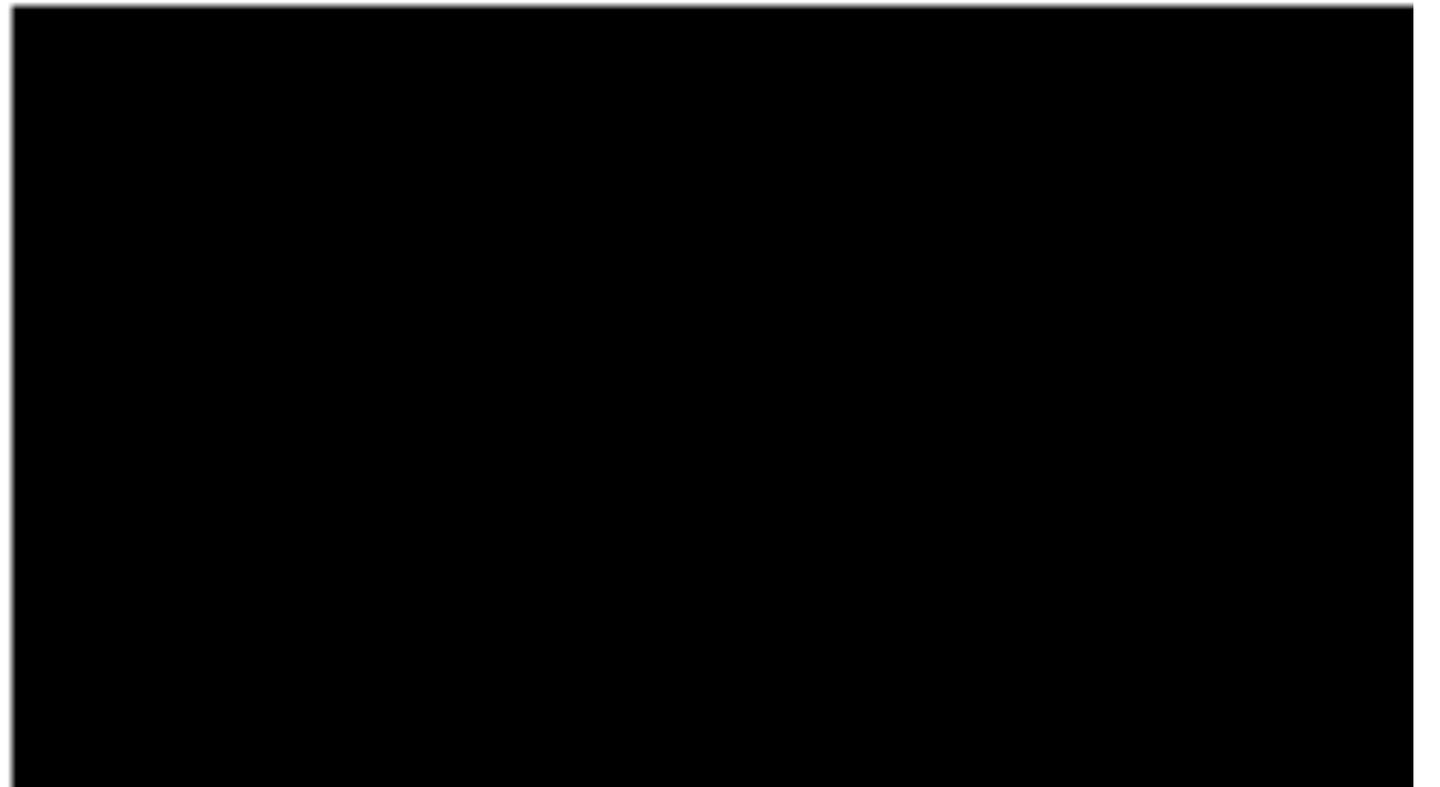
自己映像提示グループ



他者映像提示グループ



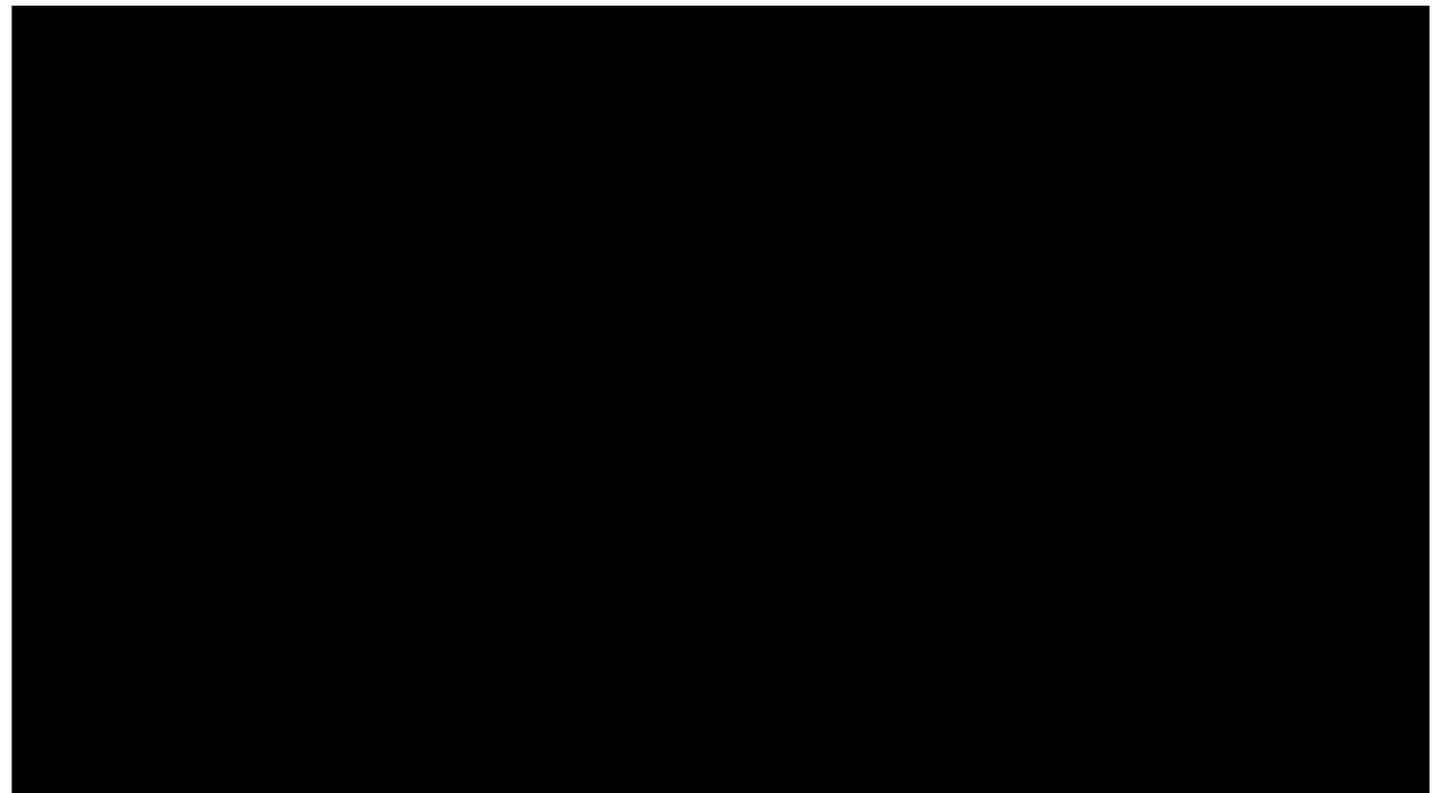
自己映像提示グループ

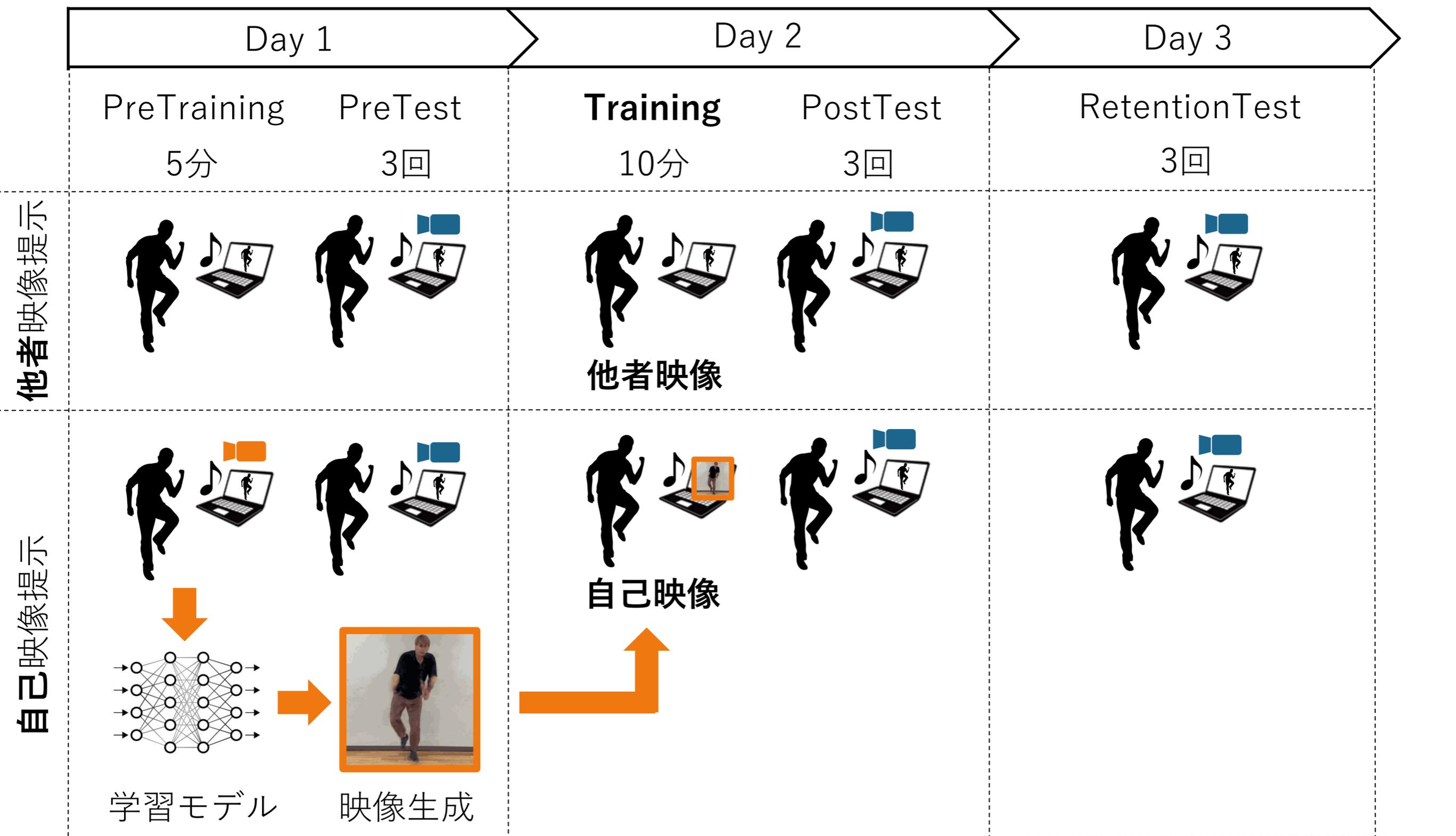


他者映像提示グループ

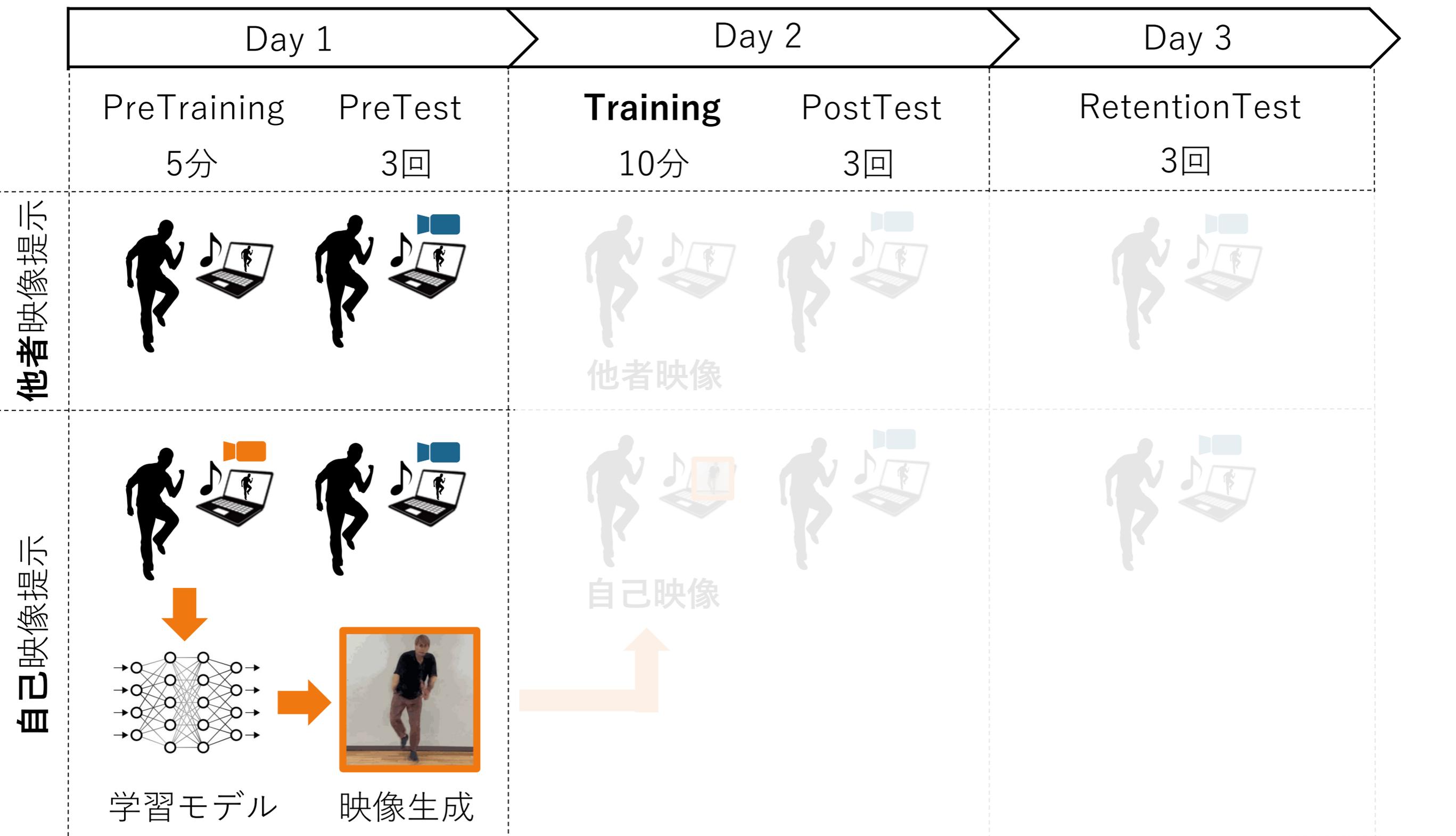


自己映像提示グループ

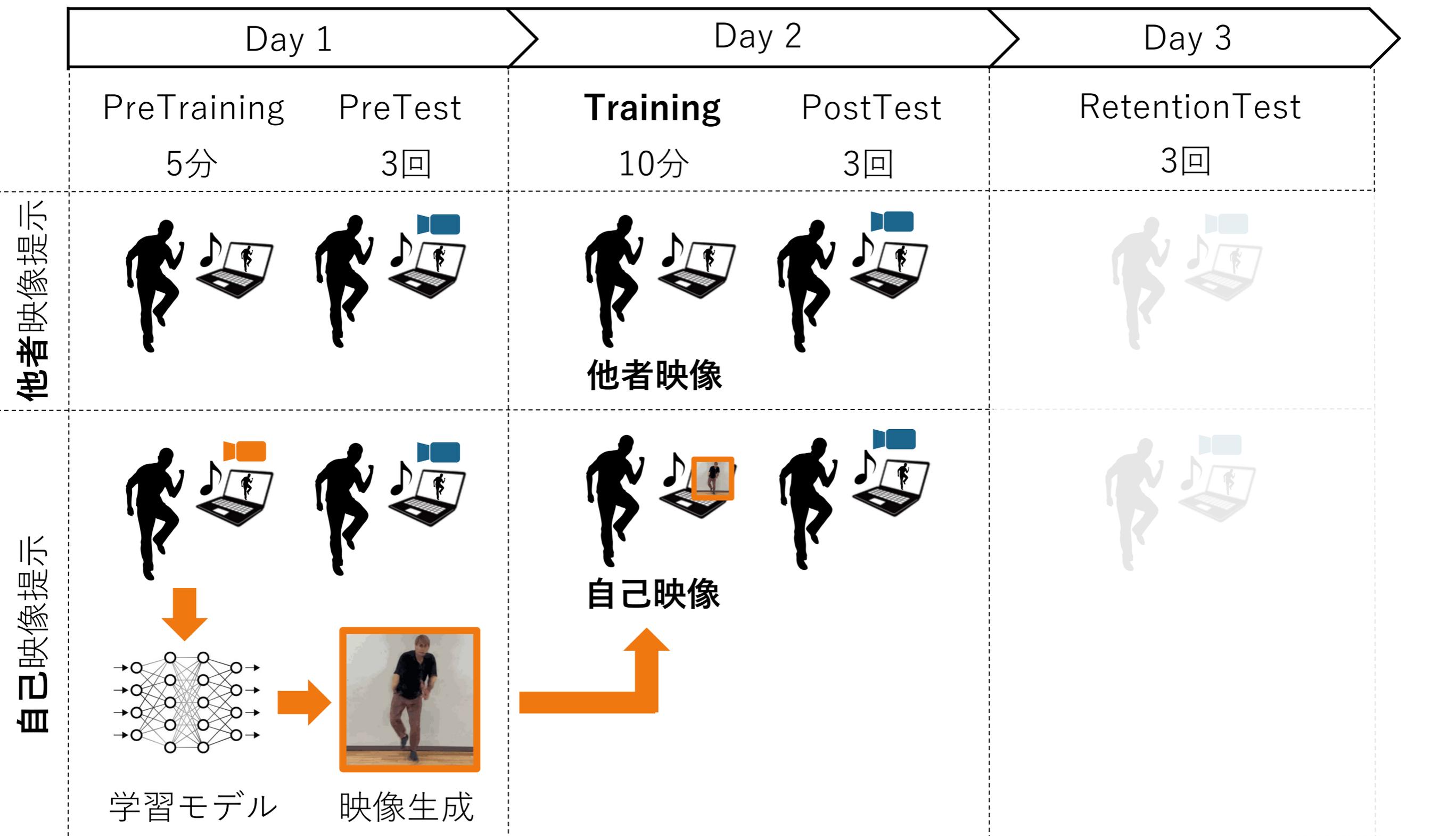




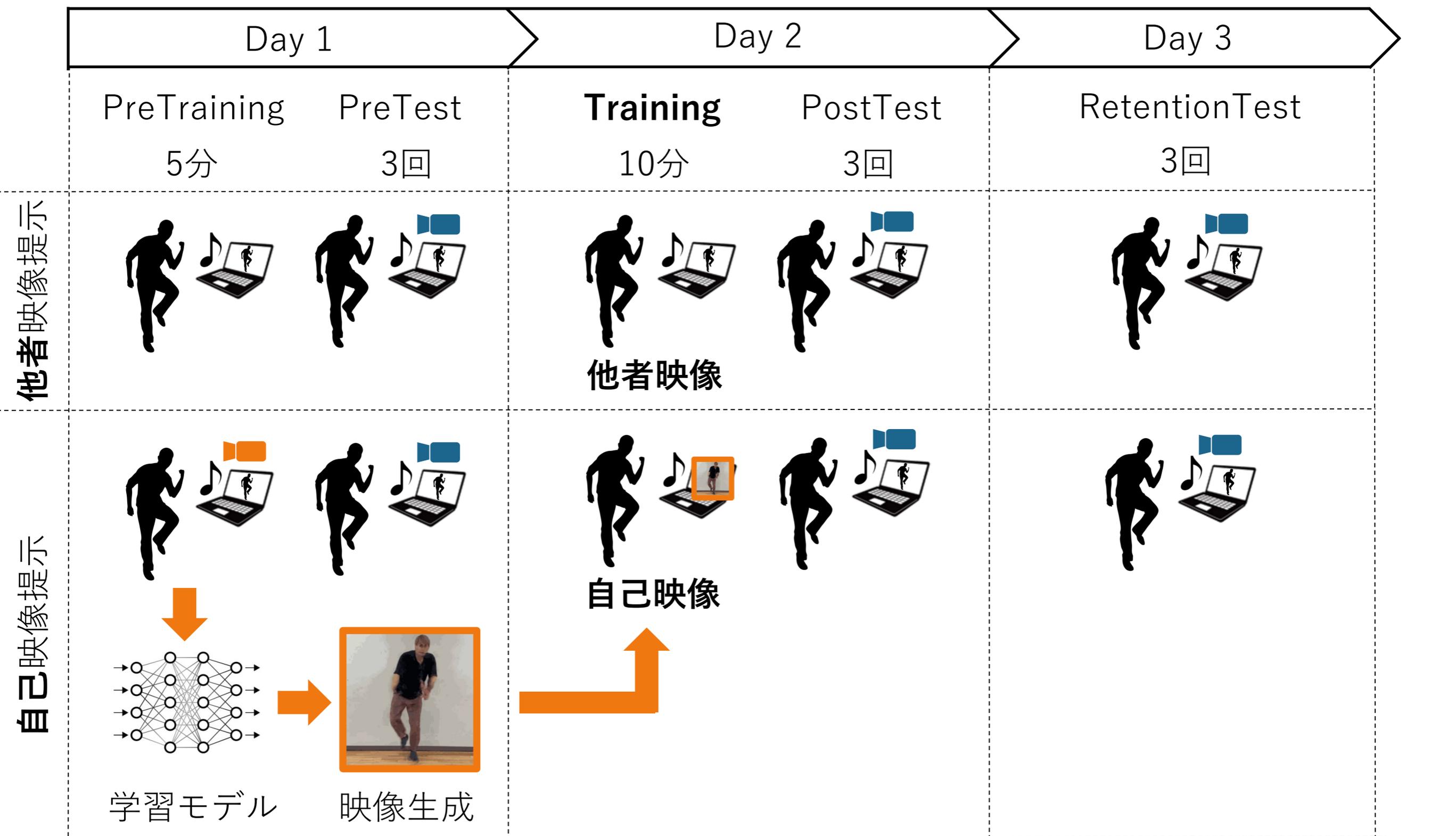
Day 1~3の映像を基にスコアリング



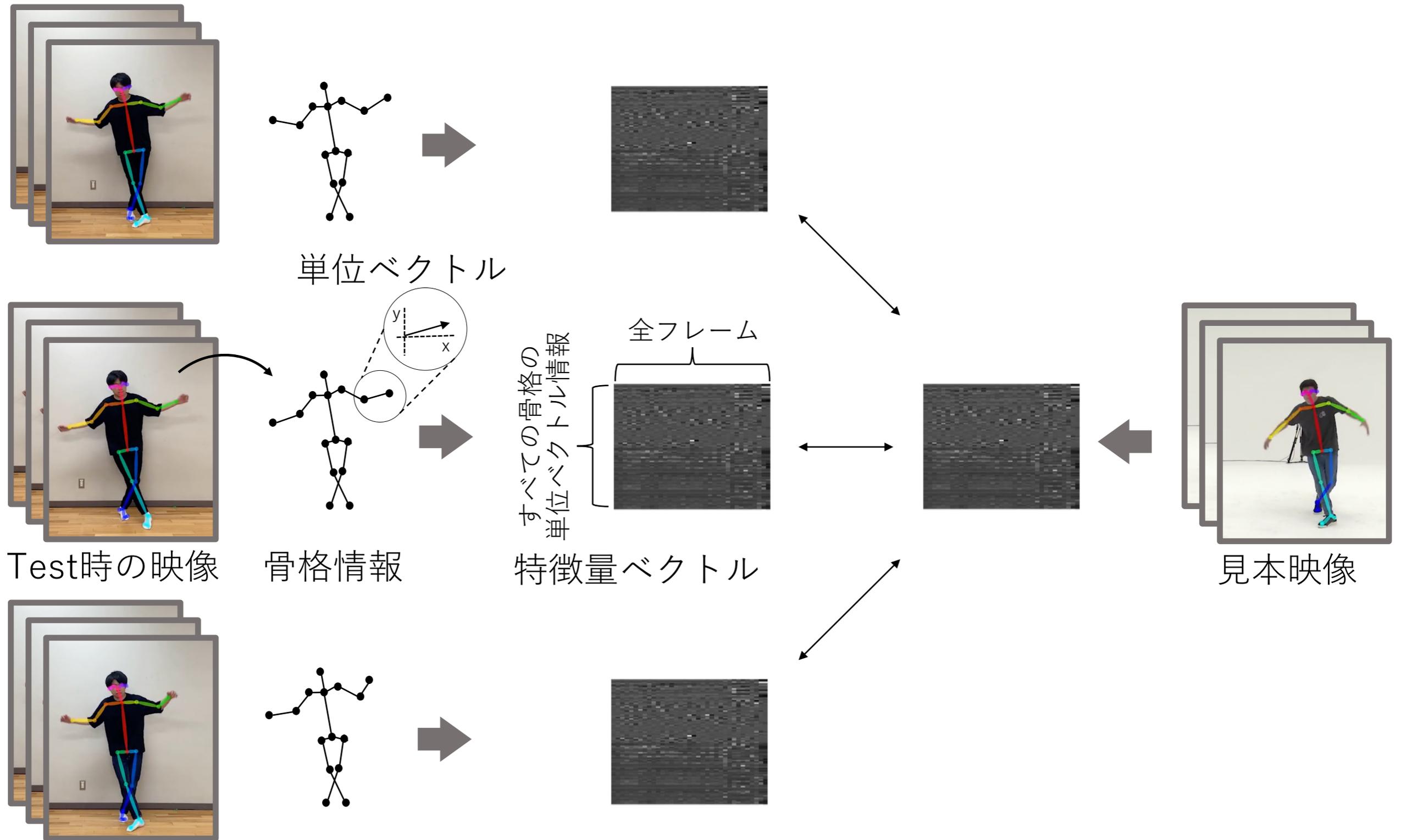
Day 1~3の映像を基にスコアリング



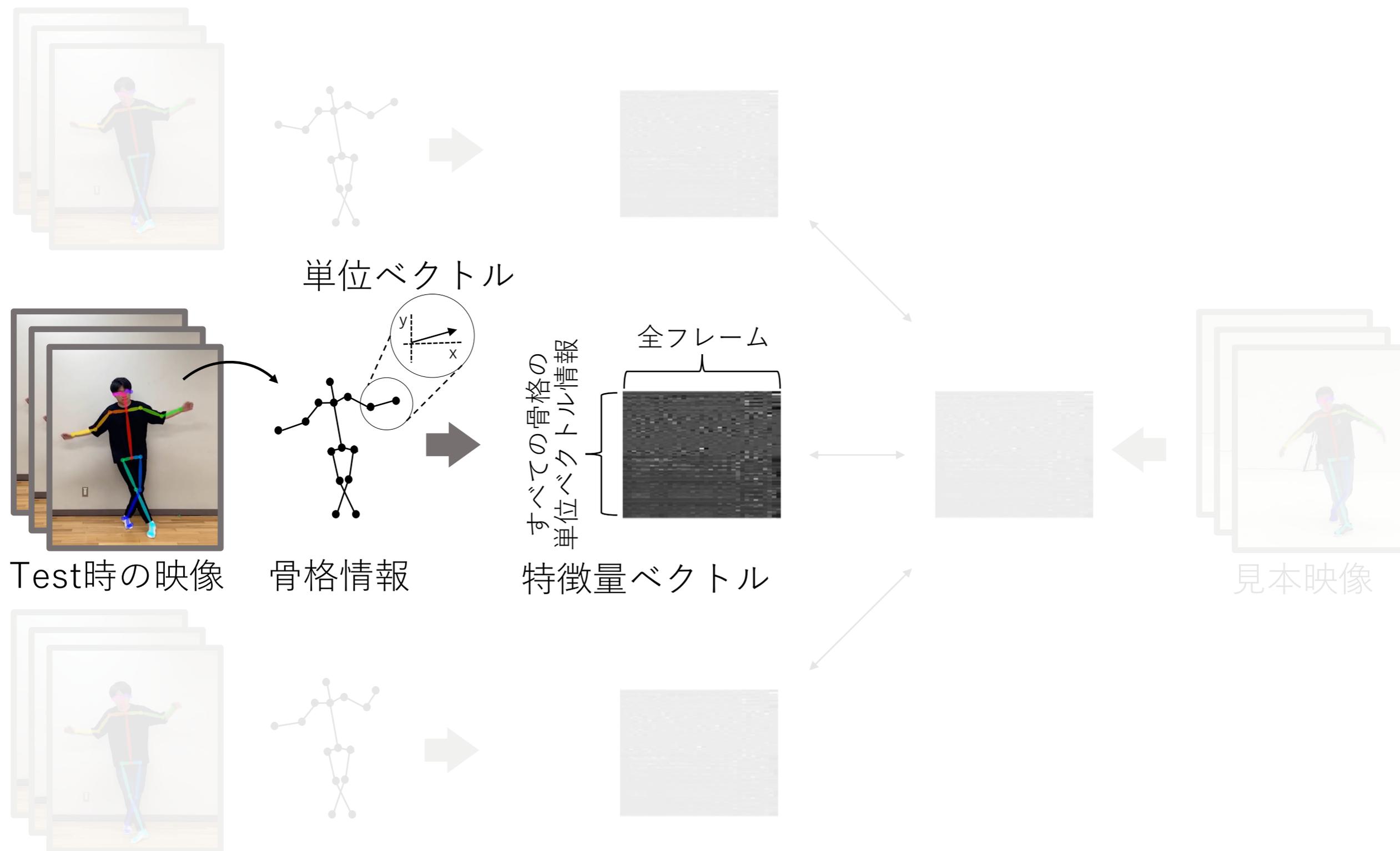
Day 1~3の映像を基にスコアリング



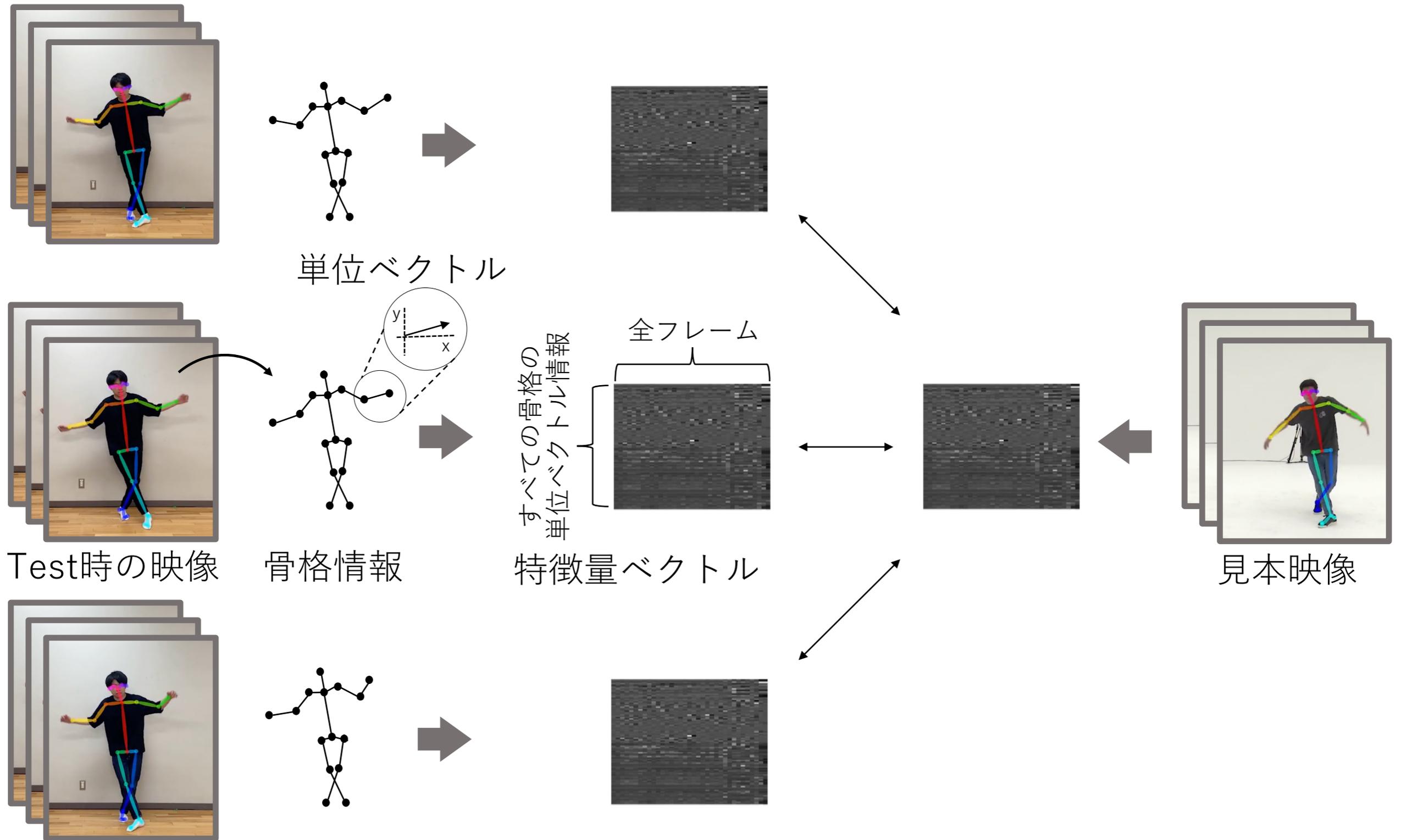
Day 1~3の映像を基にスコアリング



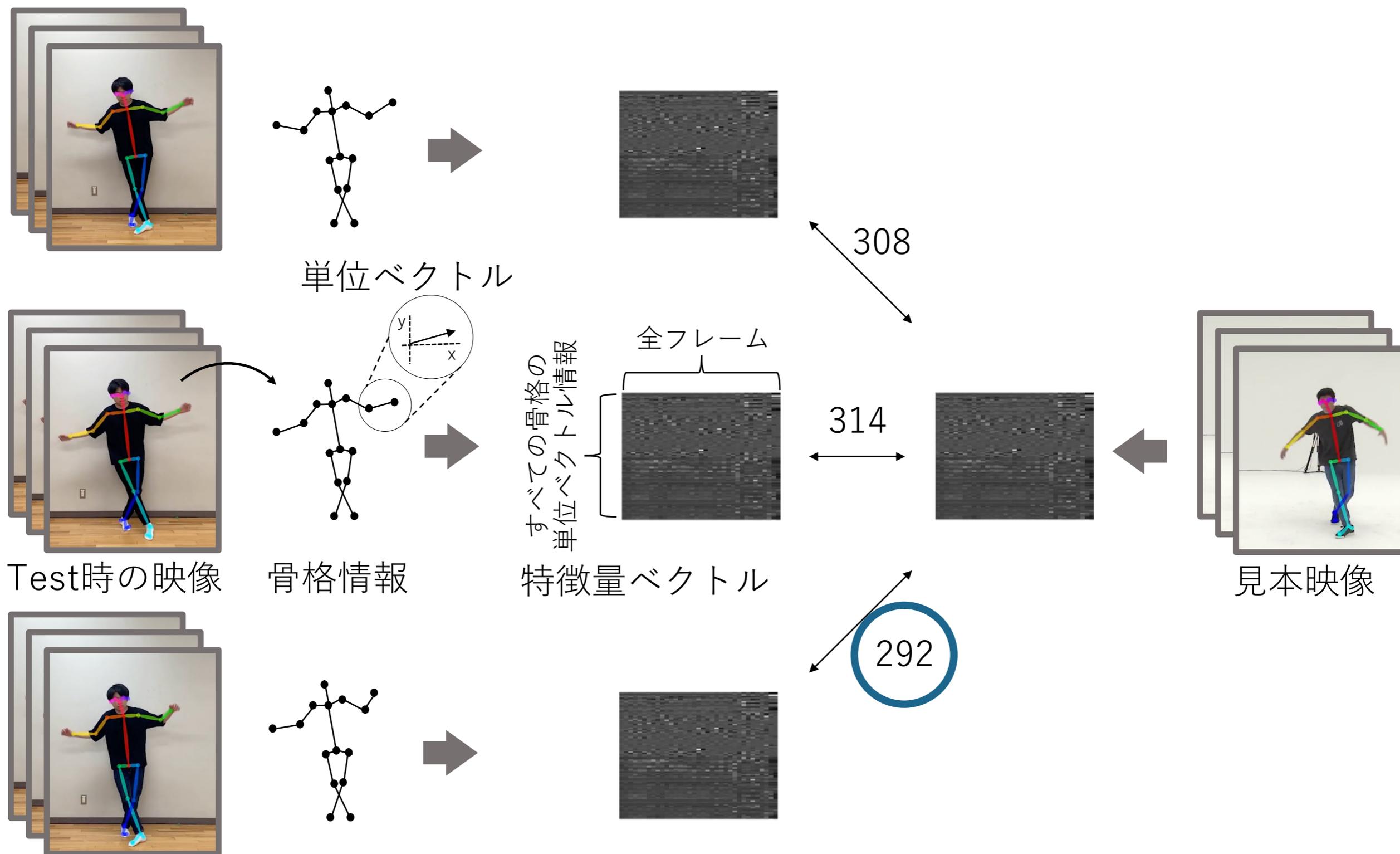
練習映像と見本映像のDTW距離をスコアとした



練習映像と見本映像のDTW距離をスコアとした



練習映像と見本映像のDTW距離をスコアとした

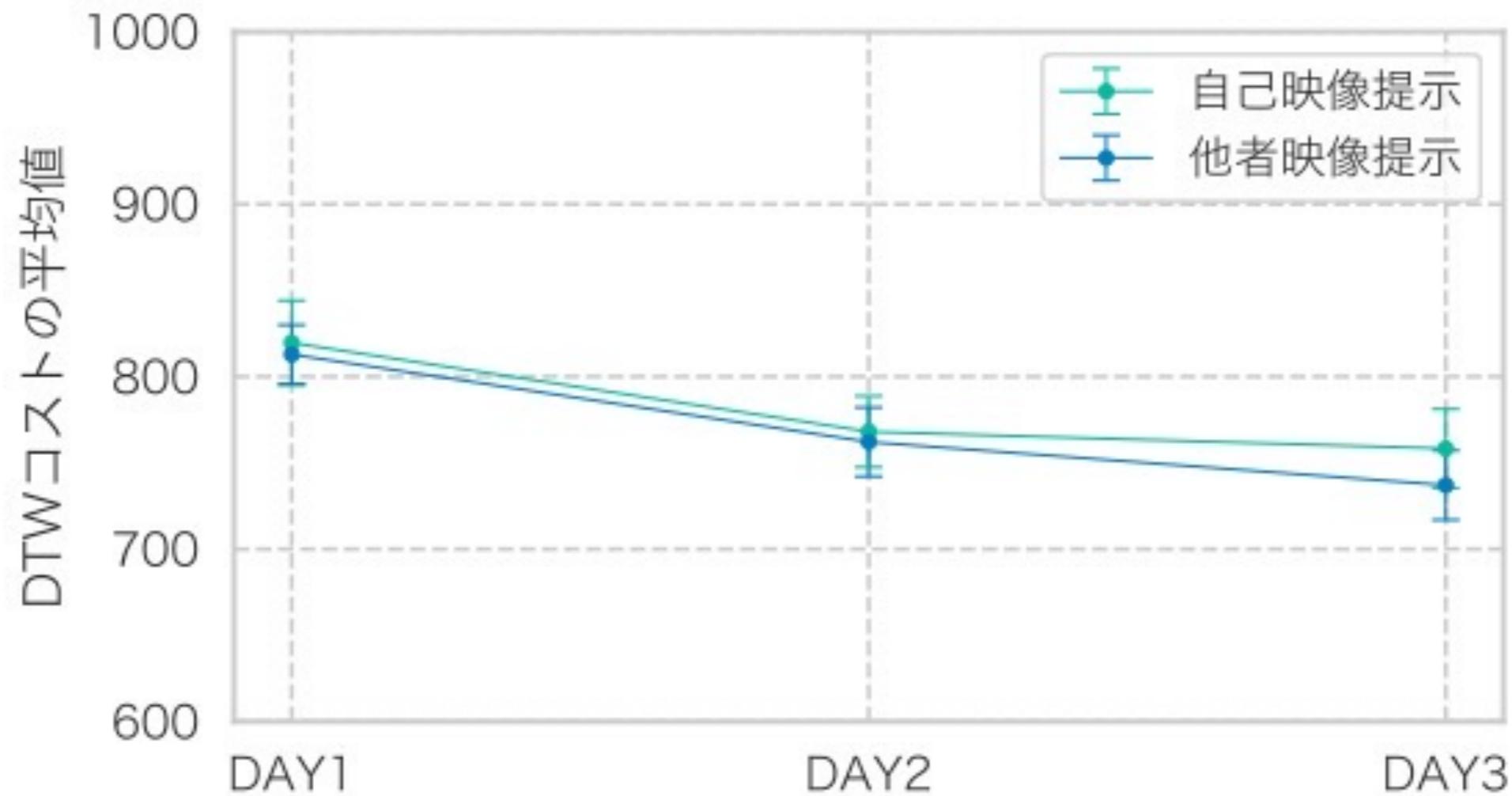


練習映像と見本映像のDTW距離をスコアとした

- ダンス動作の習得難易度について
1:非常に簡単 - 7:非常に難しい
- ダンスを習得できた
1:全く同意できない - 7:非常に同意できる
- 今後練習を続ければダンスを習得できると思う
1:全く同意できない - 7:非常に同意できる
- 自分が踊っているように感じた（自己映像提示グループのみ）
1:全く同意できない - 5:非常に同意できる

実験結果

Dance1
DAY1 - DAY2 - DAY3

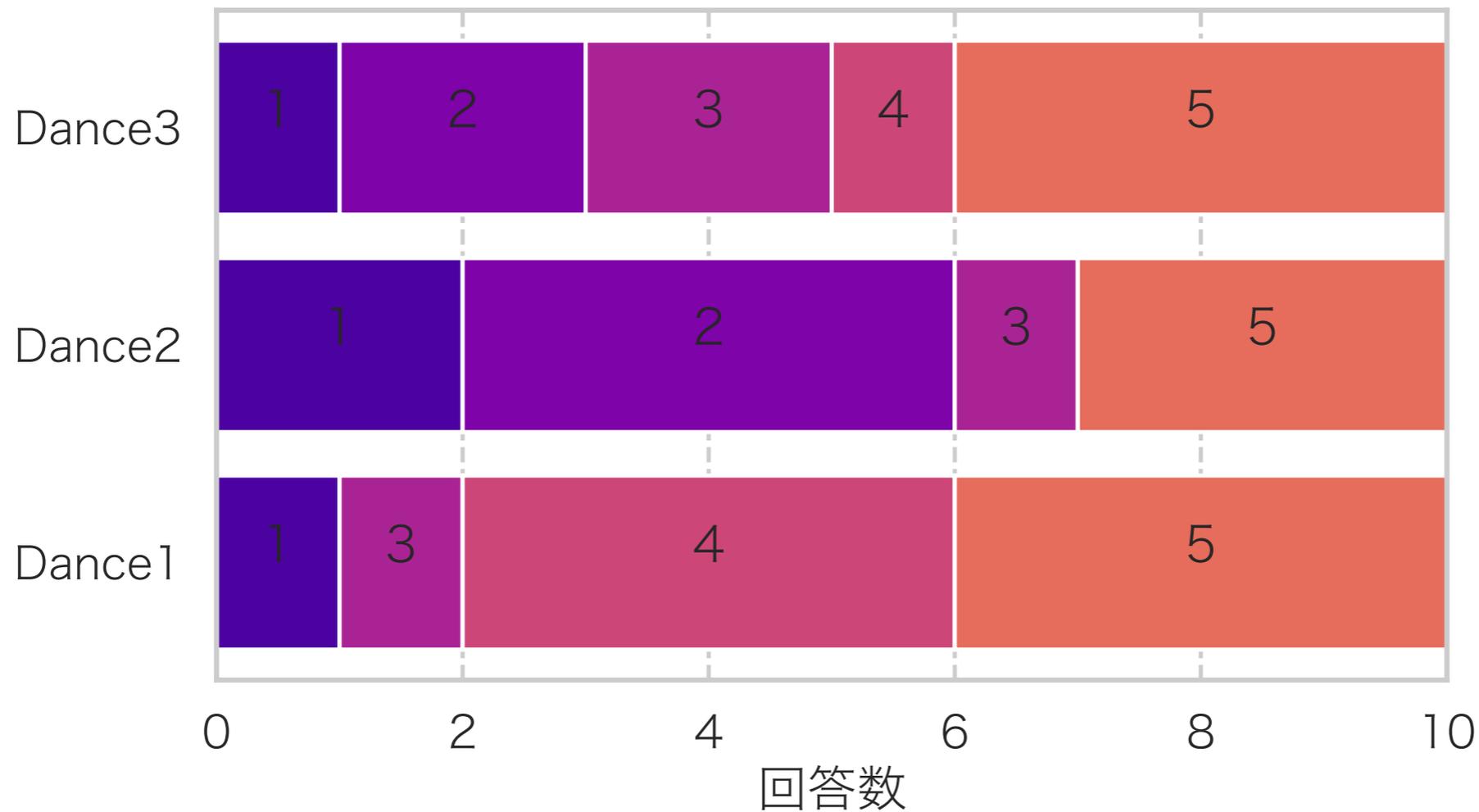


提示グループ間に有意差は見られなかった

実験結果: アンケート①

「自分がダンスしているように感じた」

全く同意できない 1 - 2 - 3 - 4 - 5 非常に同意できる

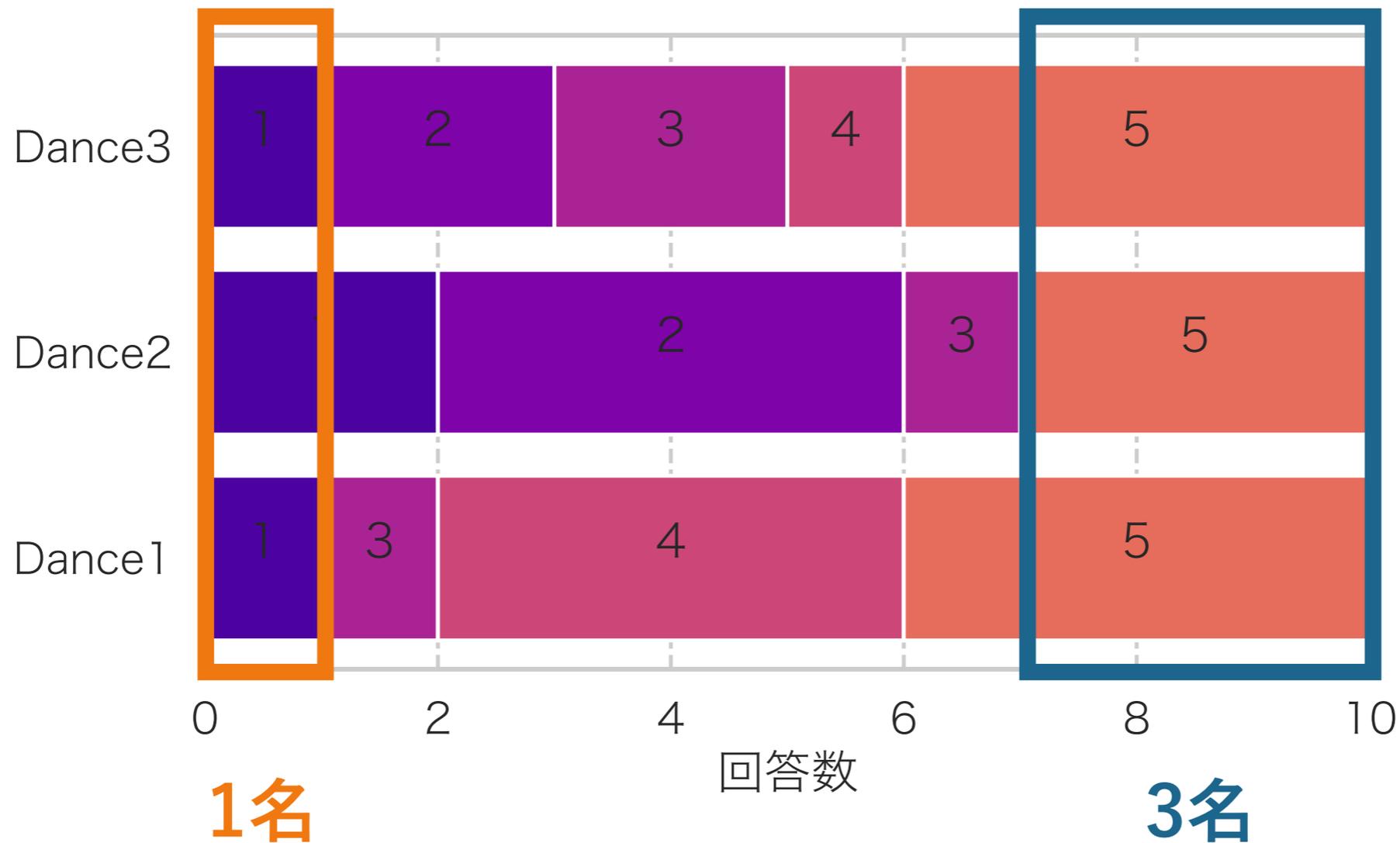


回答にばらつきがみられた

実験結果: アンケート①

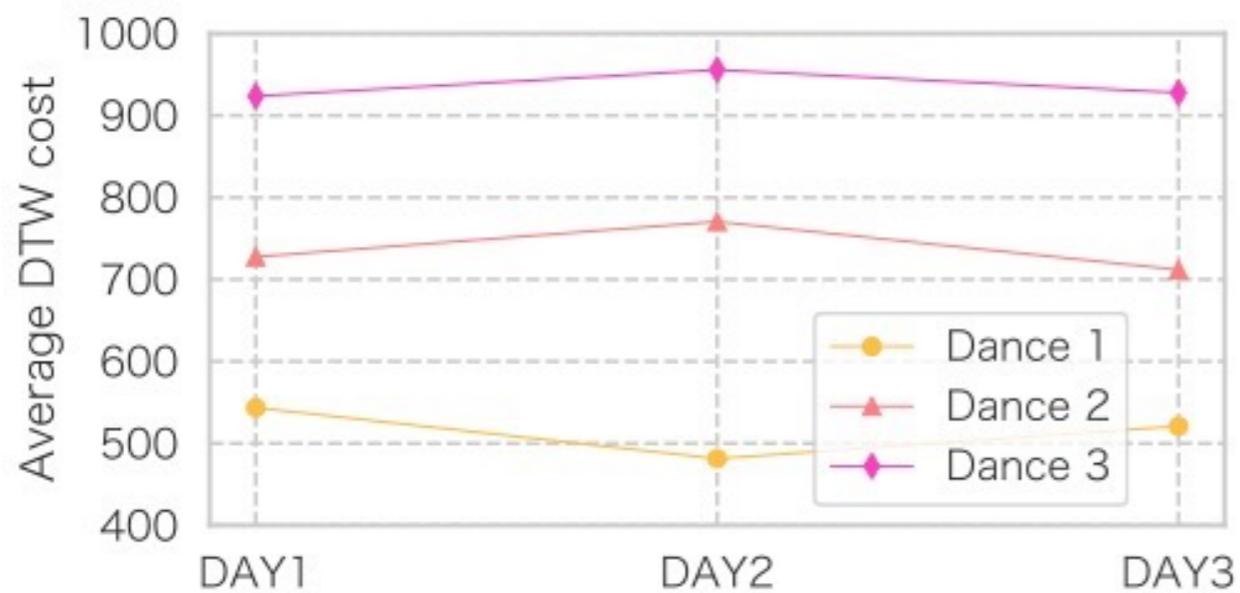
「自分がダンスしているように感じた」

全く同意できない 1 - 2 - 3 - 4 - 5 非常に同意できる

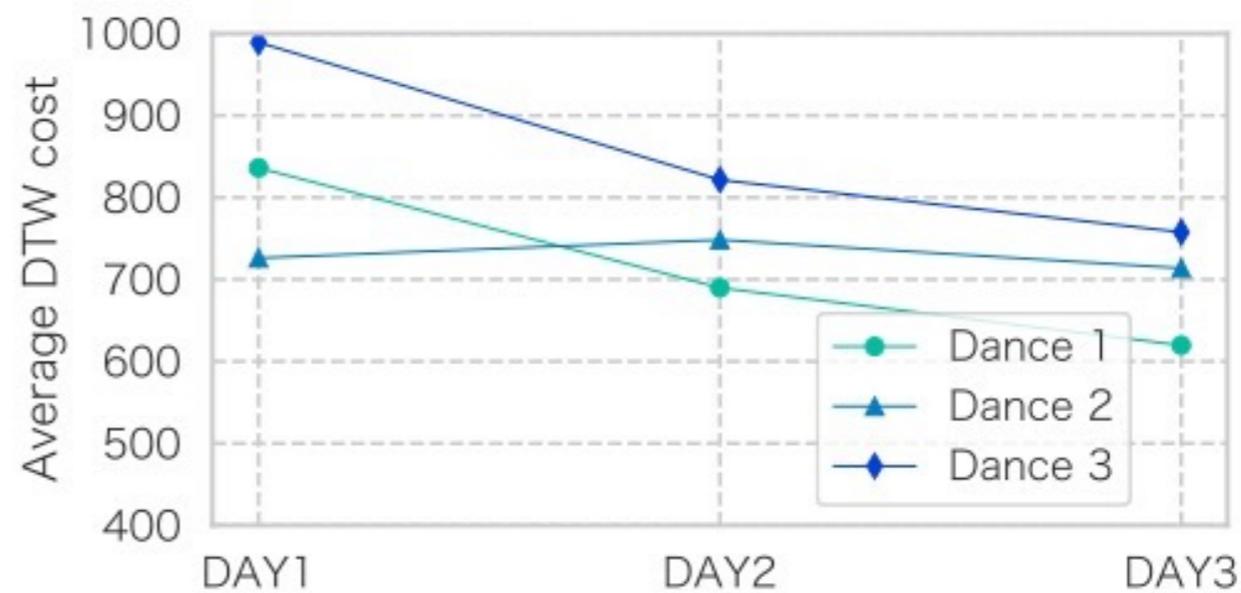


回答にばらつきがみられた

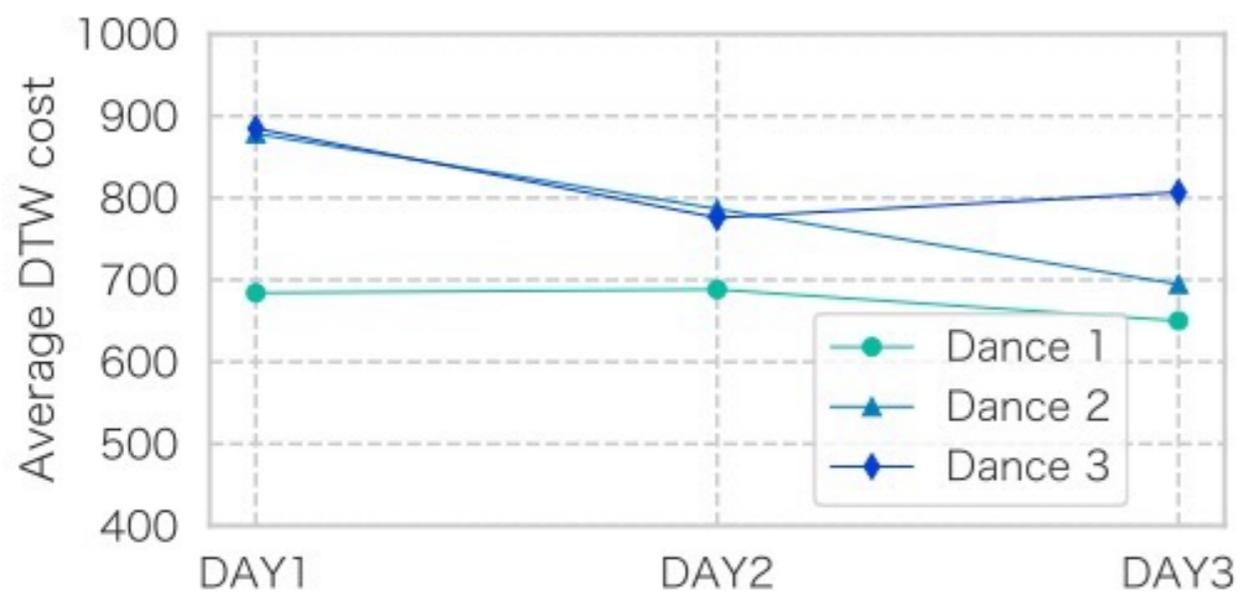
実験結果: アンケートベースのDTW距離の推移



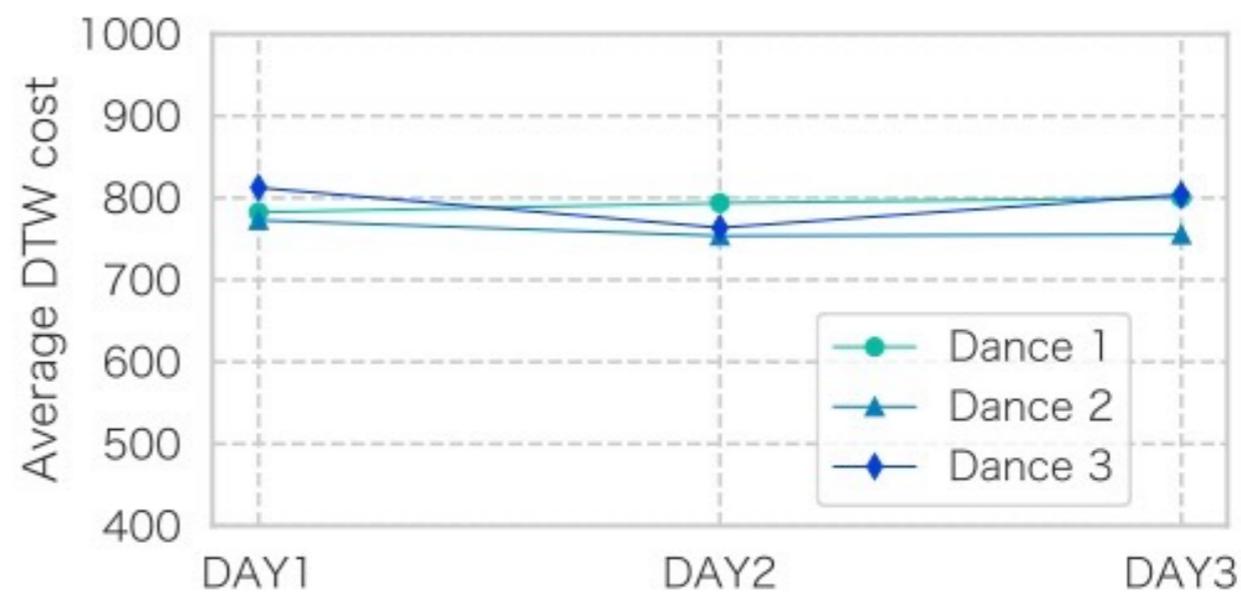
Participant A who answered 1 for all dances.



Participant B who answered 5 for all dances.



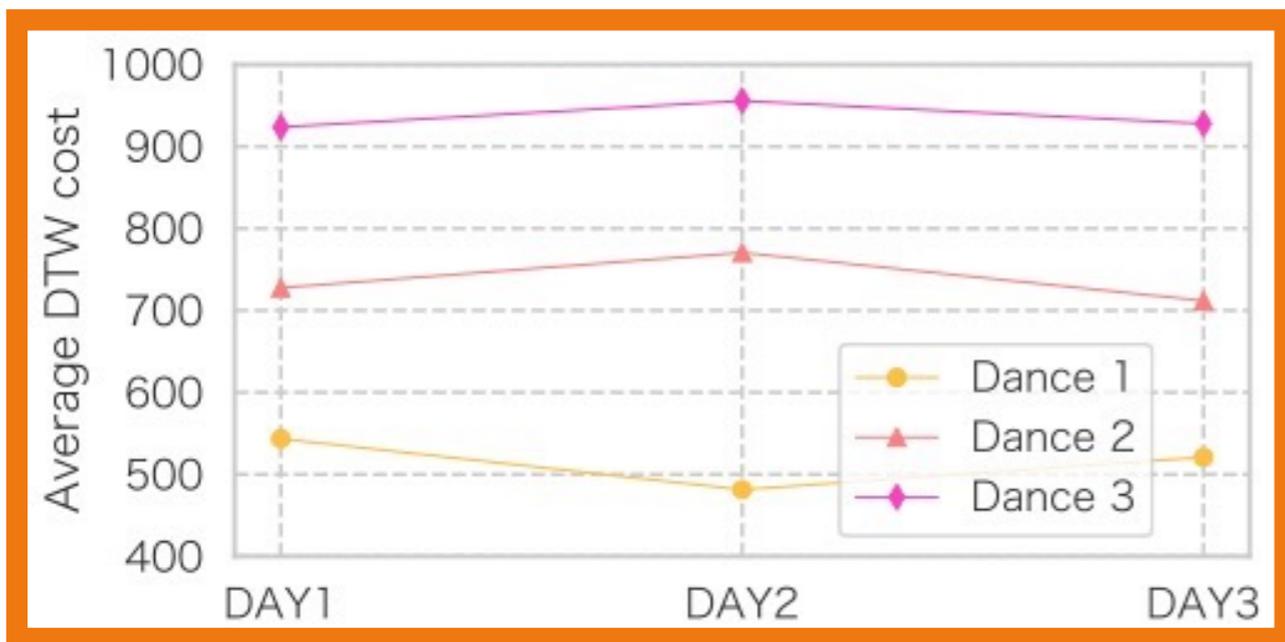
Participant C who answered 5 for all dances.



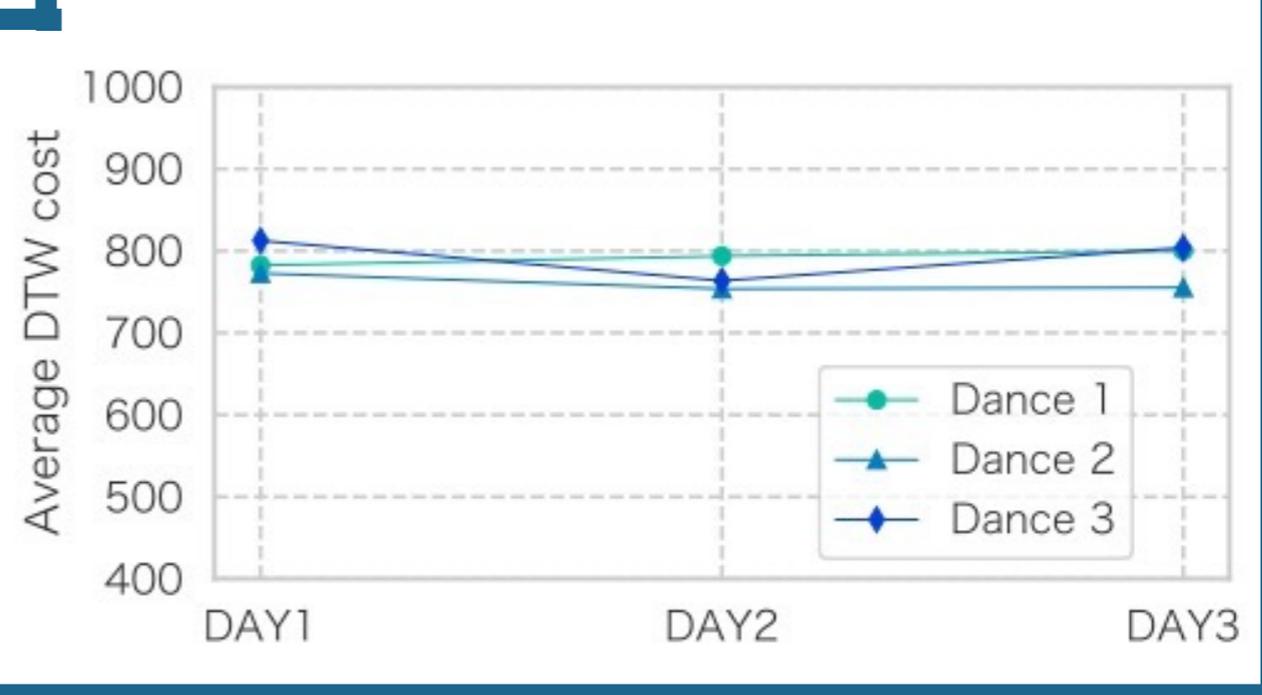
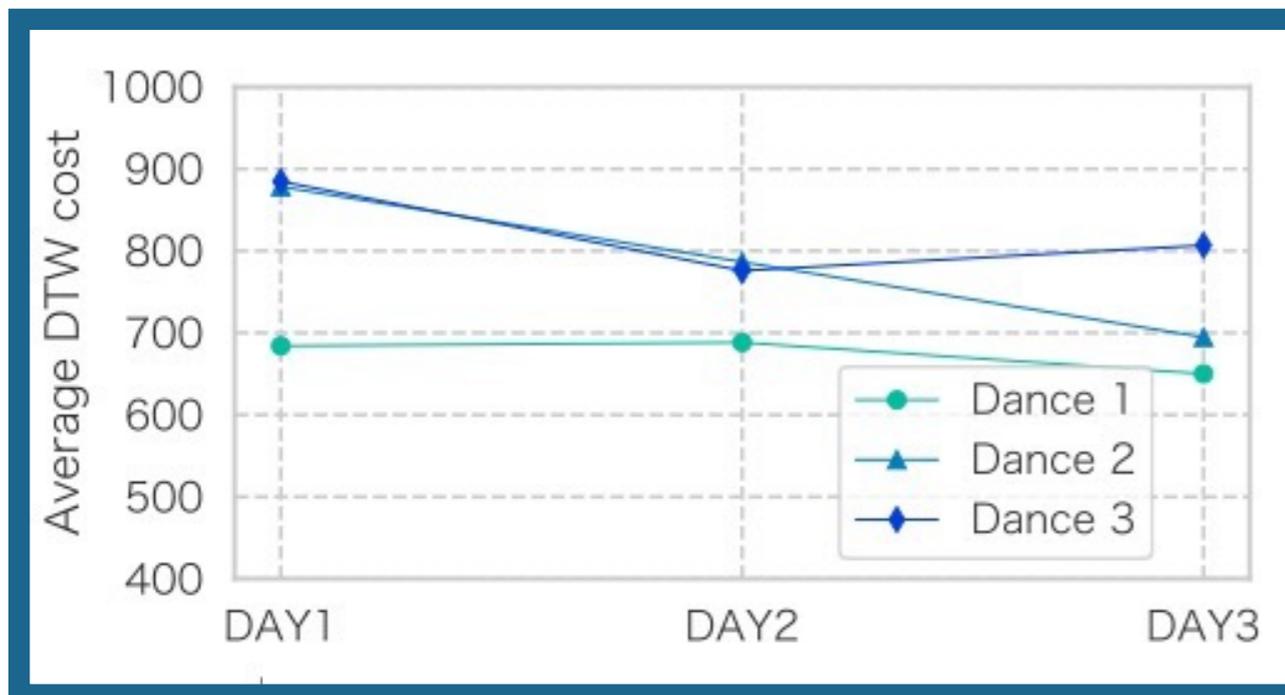
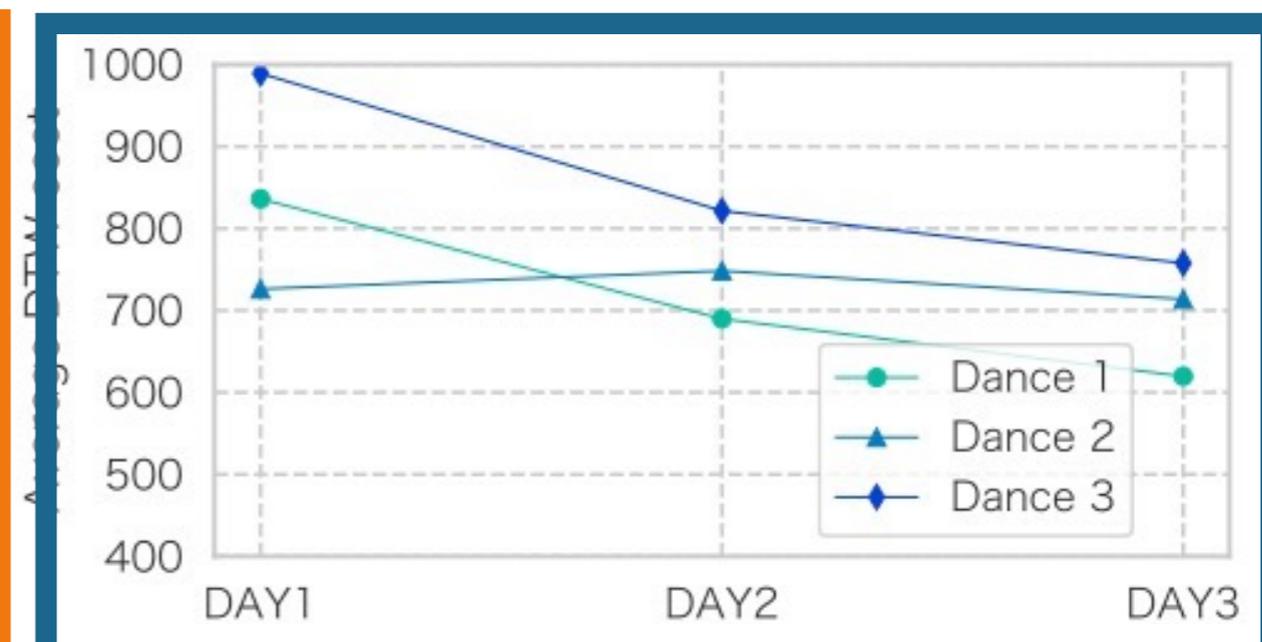
Participant D who answered 5 for all dances.

実験結果: アンケートベースのDTW距離の推移

自分がダンスしているように感じていない



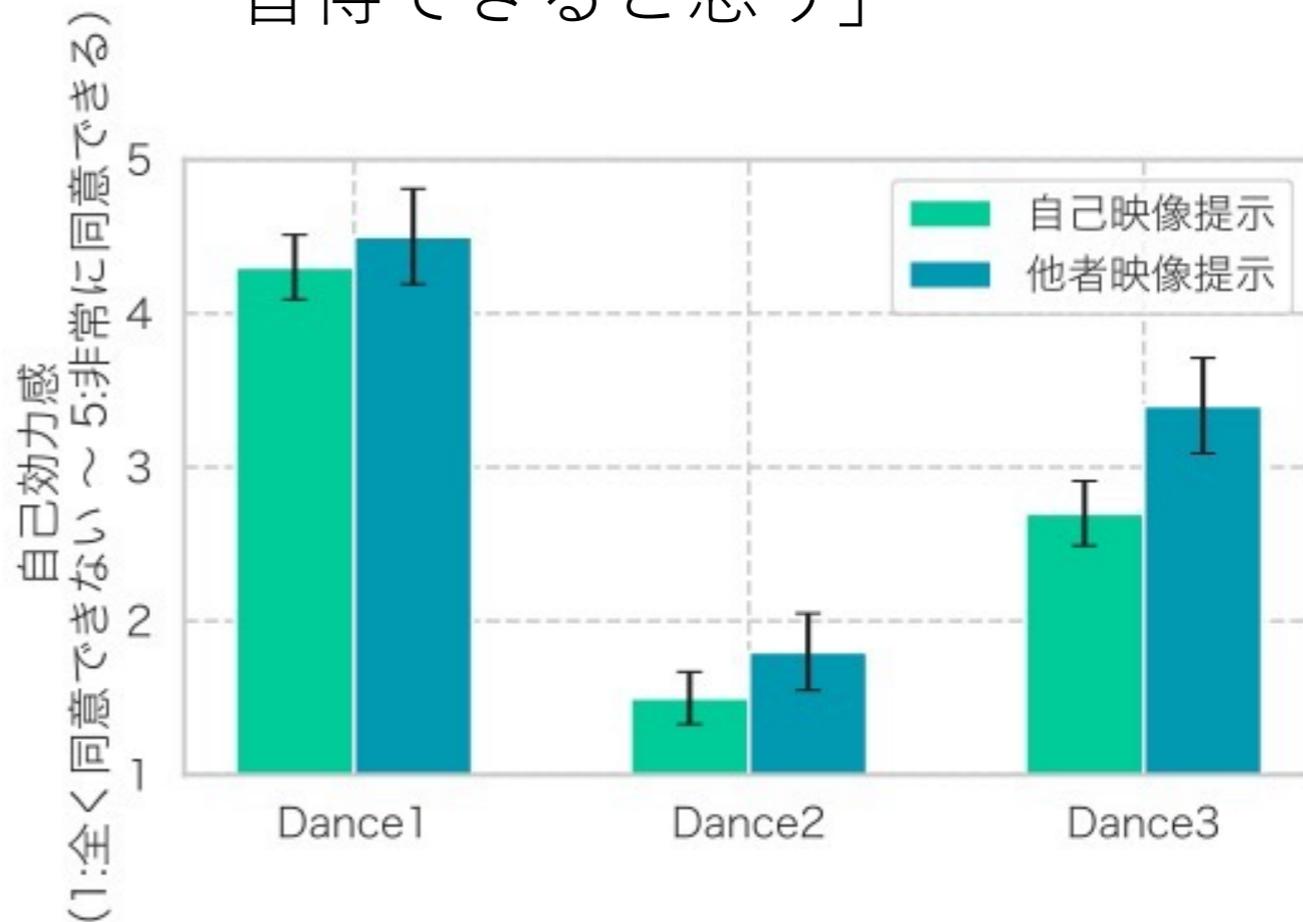
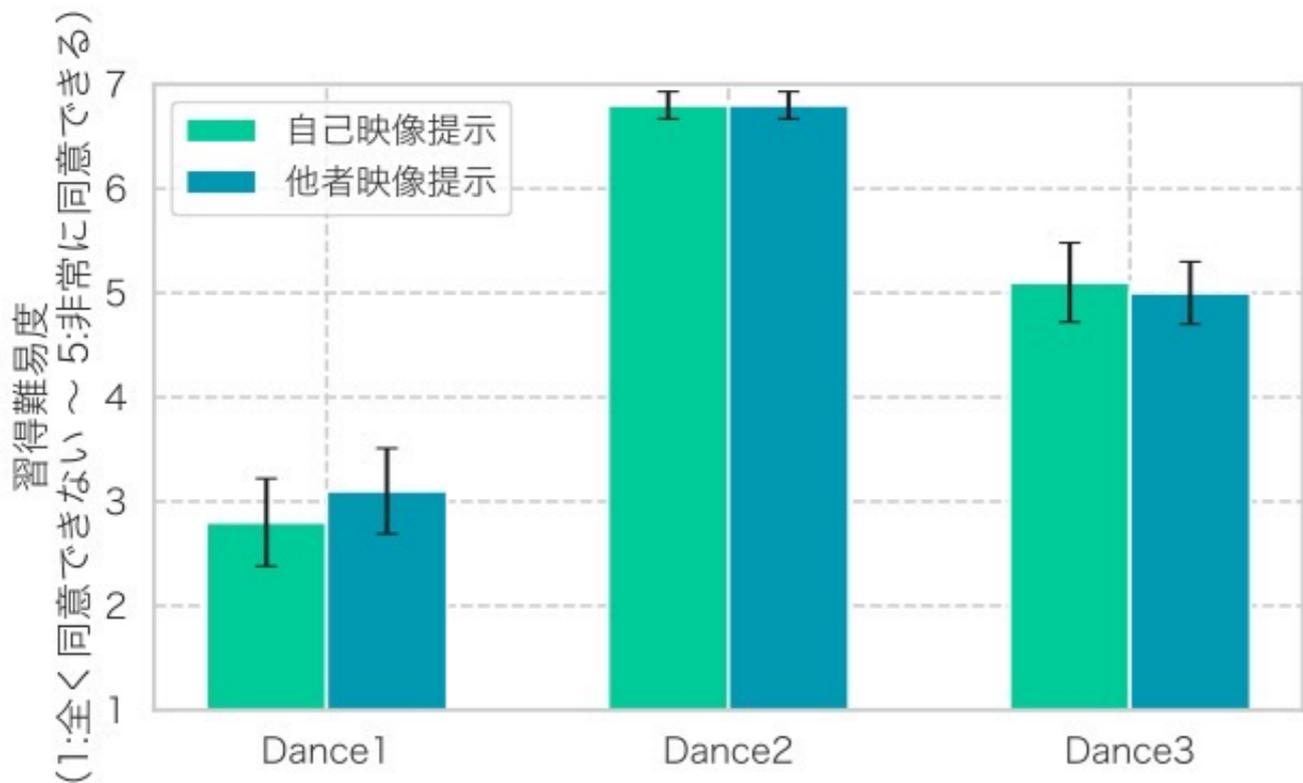
自分がダンスしているように感じている



自分がダンスしているように感じている参加者の方が
自己映像による学習が進みやすい可能性がある

「ダンスの習得難易度について」

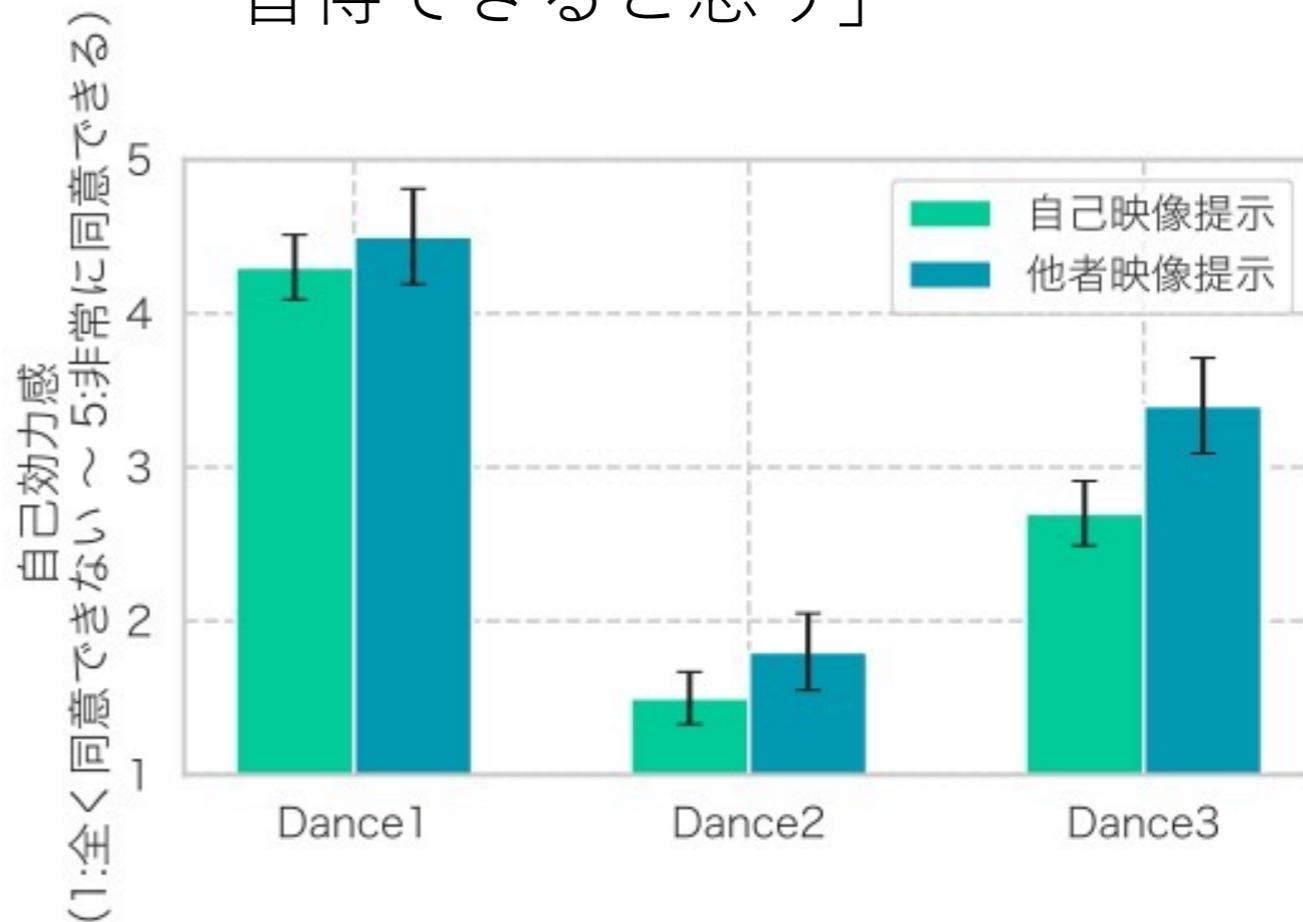
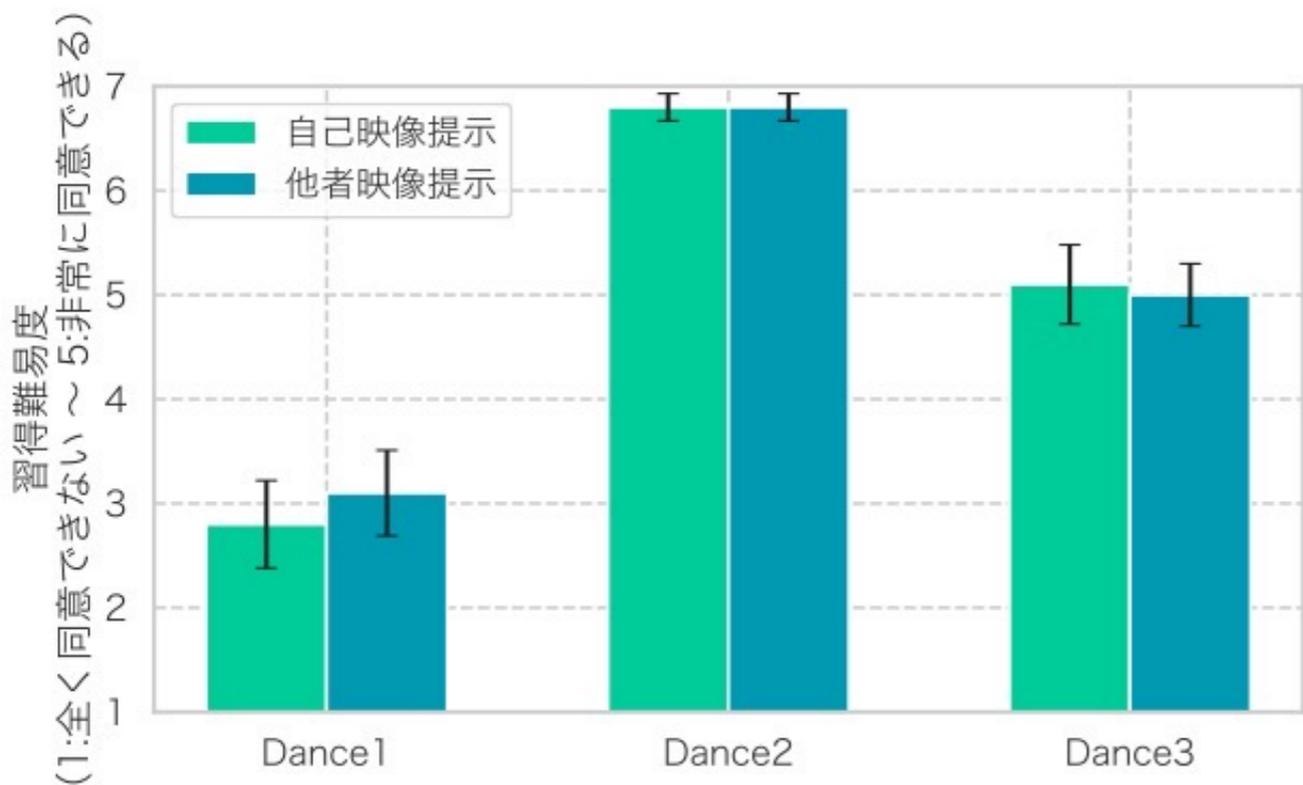
「今後練習を続ければダンスを習得できると思う」



自己映像提示グループの方が自己効力感が低い傾向がみられる

「ダンスの習得難易度について」

「今後練習を続ければダンスを習得できると思う」



自己映像提示グループの方が自己効力感が低い傾向がみられる

考察: 習得したい動作が自身のスキルレベルより離れすぎると自己効力感が低下する可能性がある

- 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。
- Pre-training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

ディープフェイクで生成された映像が動きの理解を深めるのに役立っている

- 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。違和感は特になかった
- 自分が上手く動けたらこんなふうに見えるのか、というイメージが付きやすい

動作のイメージをサポートできる可能性

- 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った
- 上手く踊れている自分が映像で見えるので、上手く踊れていると錯覚して楽しく練習できる

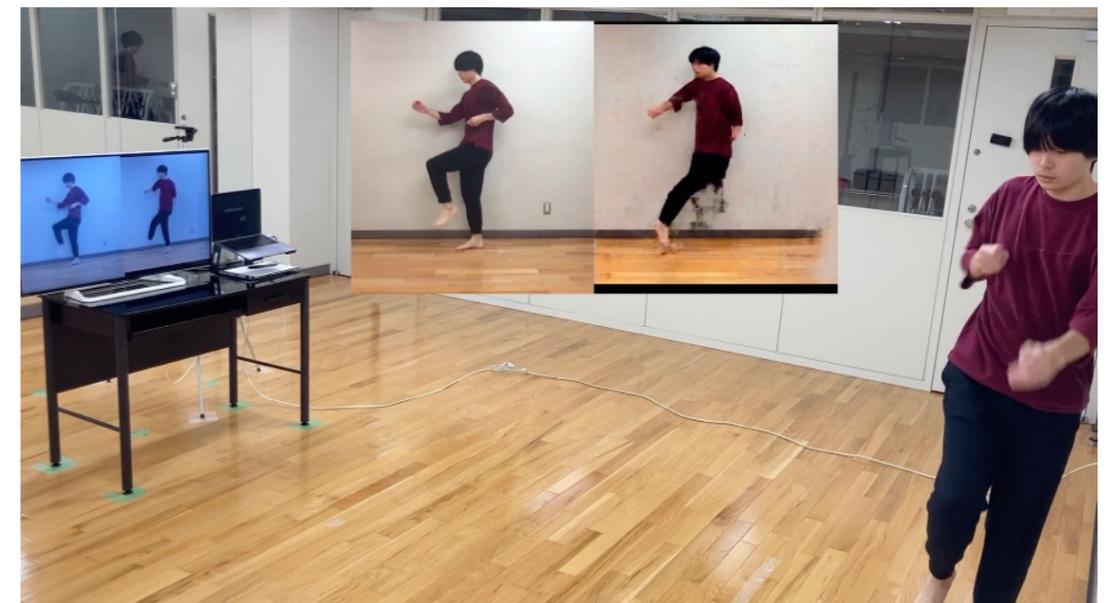
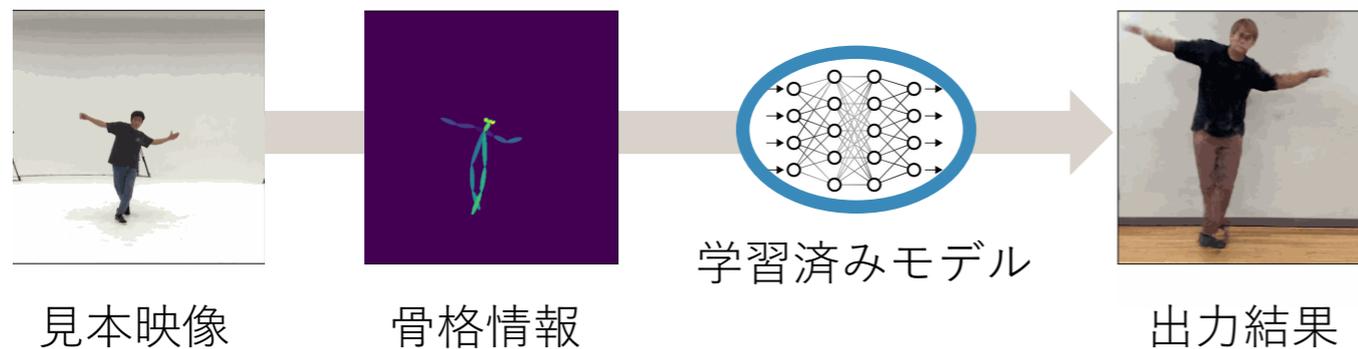
面白さという点で学習者のモチベーションを高められる可能性

- 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
- 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動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

生成の質の低さに関する指摘

- 参加者のコメントに見られるように、生成された映像の質の低さが参加者の学習に影響を与えた可能性がある。
→ **映像の不自然さを取り除いた、より高画質な映像を生成するシステムが求められる。**
- ダンスの種類ごとでグループ間に有意差は見られなかった。
- **ダンス難易度については、さらなる調査が必要。**
- 画面サイズが小さい（90cm×50cm）ため、鏡としての役割が薄れ、セルフモデリングの効果が得られなかった可能性がある。
→ **比較的大きなディスプレイを使用することで、エラー率を減らせる可能性がある。**

- 深層学習を用いて見本映像中のダンサーの動作と同じ動作を行う自己映像を生成・提示する学習手法を提案.
- 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.
- 実験の結果，提示グループ間に有意な差は見られなかった.
- アンケート結果をもとに自己映像提示の課題を整理.



スキルモーフィング

100% 初心者



スキルギャップ



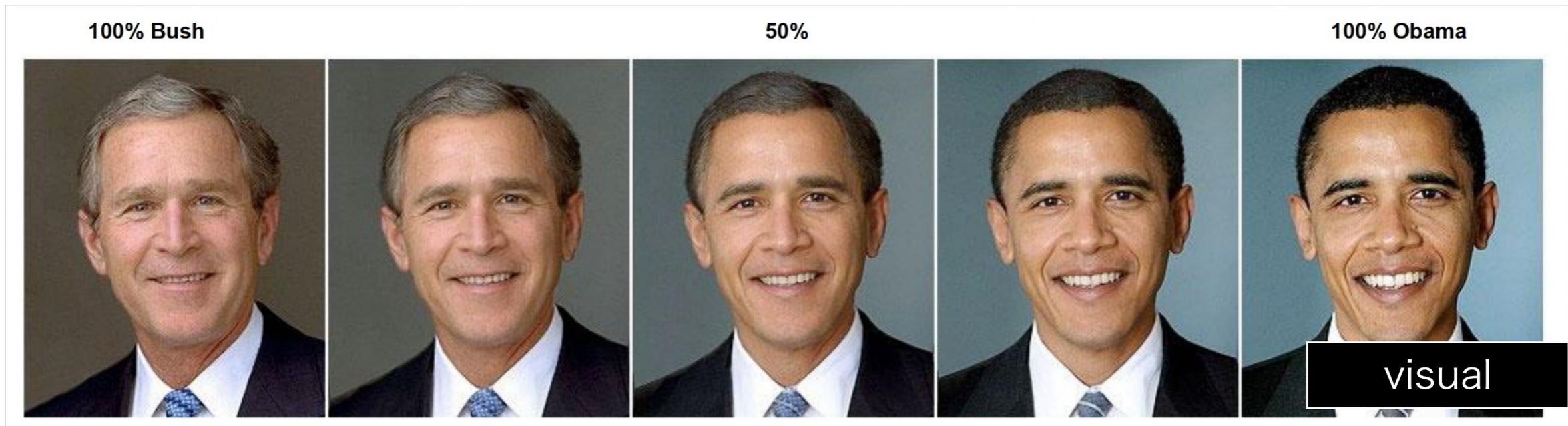
100% 上級者



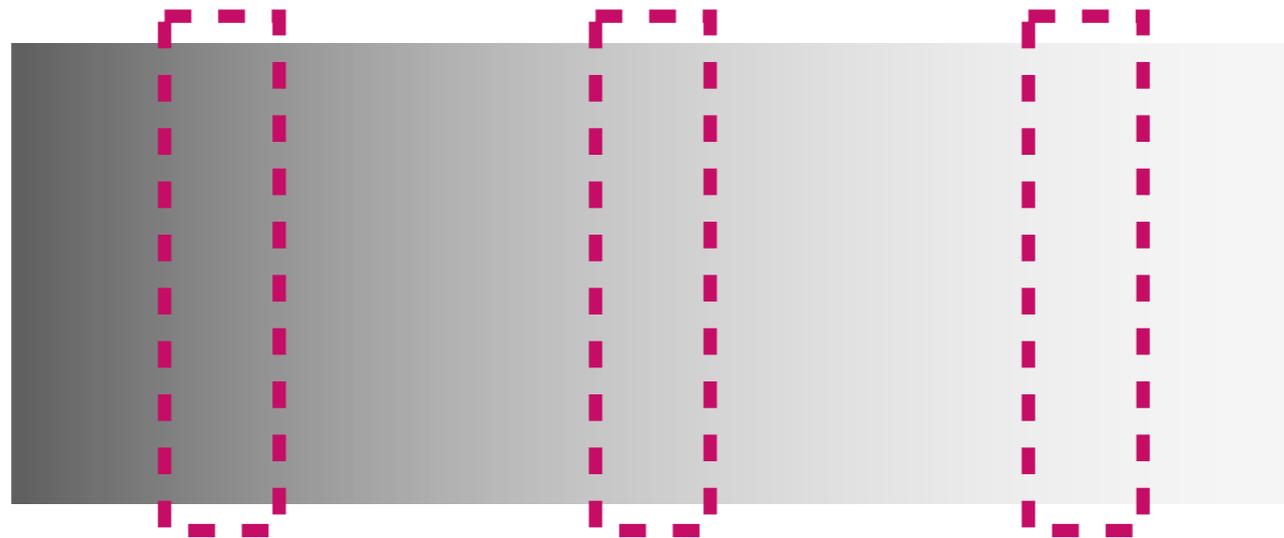
スキルモーフィング

初心者と上級者の中間レベルのダンス動作を生成する技術

自分のダンスレベルより一歩先のダンスの動きを参考にしながら練習できる



100% 初心者



100% 上級者

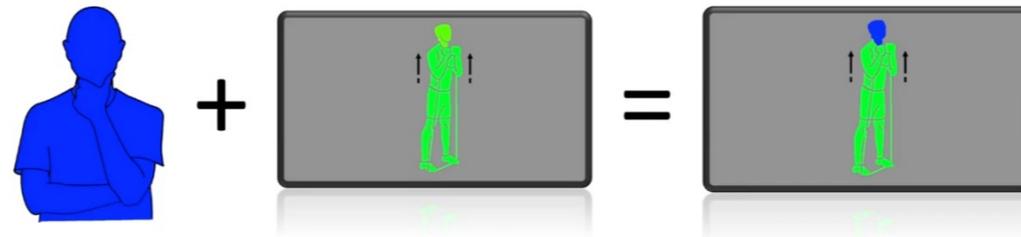


motion

FakeForward: Using Deepfake Technology for Feedforward Learning

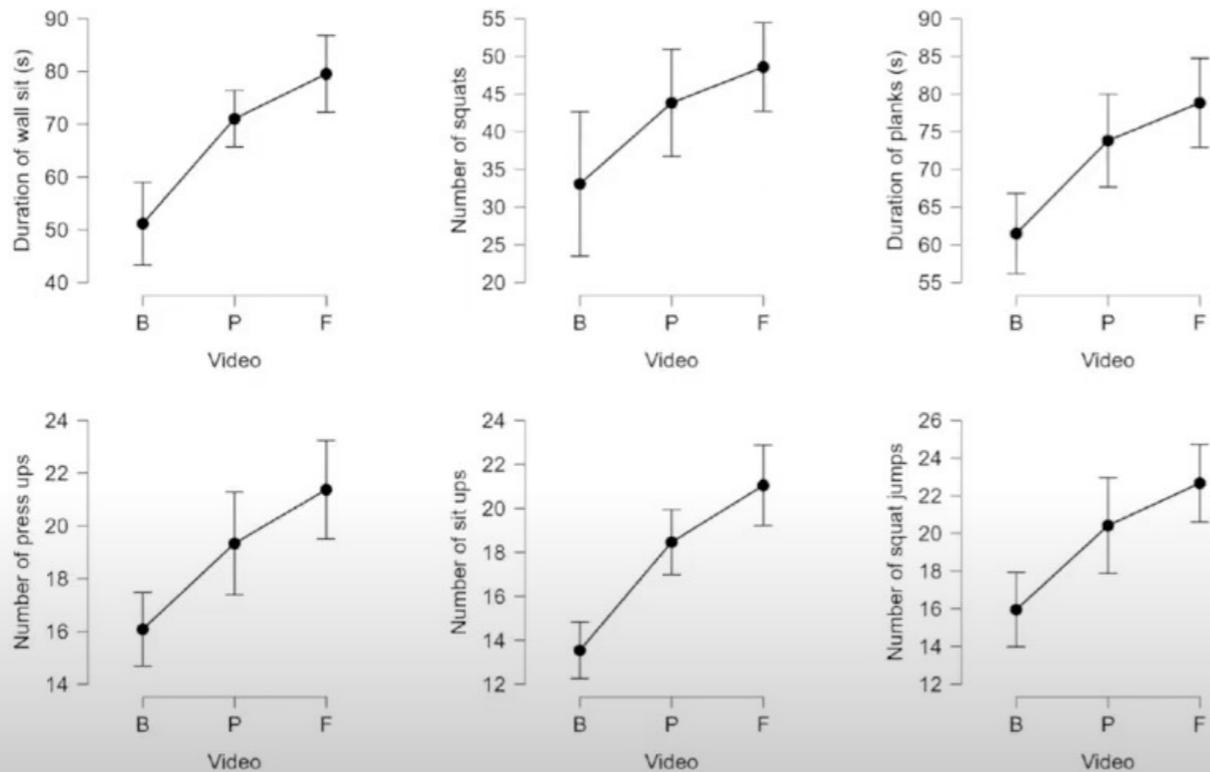
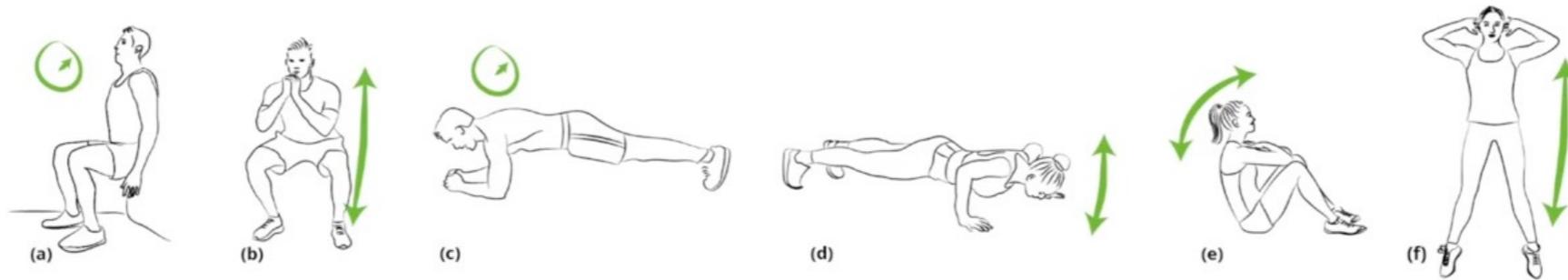
Christopher Clarke, Jingnan Xu, Ye Zhu, Karan Dharamshi, Harry McGill, Stephen Black, Christof Lutteroth (CHI '23)

FakeForward



FakeForward: Using Deepfake Technology for Feedforward Learning

Christopher Clarke, Jingnan Xu, Ye Zhu, Karan Dharamshi, Harry McGill, Stephen Black, Christof Lutteroth (CHI '23)



Measure	Test	Statistic	df	p	Effect Size
Wall sits	Student	2.134	23	0.022*	0.436
Squats	Wilcoxon	82.000	23	0.046*	0.406
Plank	Student	1.171	23	0.127	0.239
Press ups	Student	1.458	23	0.079	0.298
Sit ups	Student	2.130	23	0.022*	0.435
Squat jumps	Wilcoxon	65.000	23	0.024*	0.486

- 深層学習を用いて見本映像中のダンサーの動作と同じ動作を行う自己映像を生成・提示する学習手法を提案.
- 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.
- 実験の結果，提示グループ間に有意な差は見られなかった.
- アンケート結果をもとに自己映像提示の課題を整理.

