

招待論文 PianoSyncAR: 拡張現実で同期した手のポーズの差分を視覚化することでピアノ学習を強化

Ruofan Liu* Erwin Wu† Chen-Chieh Liao* 西岡勇人‡ 古屋晋一‡ 小池英樹*

概要. 運動の技能習得において、学習者が自身の動作と目標の動作との時間的及び空間的な差分を認識し学習することが一般的である。しかし、自由度の高い緻密な動作では、これらの差分を正しく理解し修正することが困難である。特に洗練されたスキルを持つ中上級者の場合は、より細かな修正やさらなる上達を求められるため、この問題は顕著となる。本研究では、ピアノの学習における手の動きを効率よく最適化するための拡張現実 (AR) システム「PianoSyncAR」を提案する。提案システムでは、教師の手の姿勢を学習者の手に重畳表示する。ユーザ実験では、12人のピアニストを対象としてシステムの有用性を評価する。その結果、映像ベースのスキル学習方法である従来のタブレット画面に対して提案システムにはいくつかの利点が示され、ピアノ演奏の技術向上のための補助的なツールとして AR トレーニングの可能性が示された。

1 はじめに

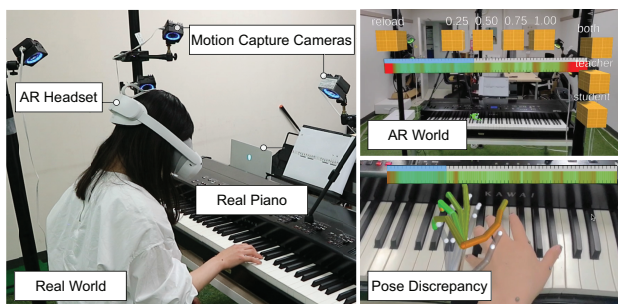


図 1. 提案システムの概要. ピアニストがピアノを演奏する際の手の姿勢と打鍵データを現実世界で収集し、ユーザとの差分を AR 環境で可視化する。

ピアノ演奏は非常に複雑な動作であり、ピアノの学習には模倣が良く用いられる。Feedback-Error-Learning 学習法 [1] によれば、技能の習得は目標と自己の動作の時空間的な不一致から学習される。しかし、高度なピアノ演奏では、自由度 (DoF) が大きく、その不一致を認識し修正することが難しい。この問題が特に深刻となるのは、経験豊富な学習者がさらに高度な技能に挑戦する場合である [2]。それは、習熟度が上がるにつれて、理想な動きとの差がより微小となり、さらなるスキル向上が難しくなるためである。

本論文では、ピアノを演奏する学習者の手の上に、時間的に変化する基準となる手の姿勢を重畳表示す

る AR トレーニングシステム「PianoSyncAR」を提案する (図 1)。提案システムでは、Keystroke-Temporal Cycle-Consistency (K-TCC) によって同期された基準となる手と学習者の手の姿勢から 3次元表現を再構成し、両者の手の姿勢の差分を AR 環境で表示する。これにより、ピアニストは没入型トレーニング環境において、姿勢の不一致から直感的に学習することができる。

2 システムデザイン

2.1 ハードウェア

提案システムでは、ピアノ演奏中のキーストロークを記録するために、先行研究 [3] でも用いられている非接触光学センサーシステムを使用する。

手の姿勢のトラッキングには、7台の OptiTrack Prime 13W カメラを 180FPS で撮影するモーションキャプチャシステムを採用する。

ユーザが AR システムを体験する際の HMD には、PICO 4 を用いる。このデバイスは一般的に VR アプリケーションで利用されているが、シースルーや AR の機能も備えている。

2.2 モーション同期

ユーザの手と参照する手の差分を認識するためには、2つの動作を同期させることが重要である。動作を同期させるため、先行研究ではタイミングに異なるビデオ入力を整列させる Temporal Cycle-Consistency (TCC) の有効性が実証されている [4]。

TCC ネットワークは一般的なビデオ位置合わせのための自己教師付き学習方法として機能するが、本論文では手の動作のシーケンスに適用するための修正されたネットワークである Keystroke-TCC (K-TCC) を導入する。88個のキーストロークを入力し

Copyright is held by the author(s). This paper is nonrefereed and non-archival.

* 東京工業大学

† 華為技術日本株式会社

‡ 株式会社ソニーコンピュータサイエンス研究所

た後、K-TCC を DTW アルゴリズムと組み合わせて活用することで、ピアノ演奏におけるユーザとお手本の両方の手の動きを正確に同期することができる [5].

2.3 AR インターフェース

2.3.1 差分の視覚化

フレーム間差分 ユーザは、視野の中央にレンダリングされた固定タイムラインを見ることができる (図 1 右下). タイムライン上には一連のエラーバーが表示される. エラーバーは、(1) 演技全体を通して不一致のあったフレームを強調表示するだけでなく、(2) 指定された閾値に基づいてエラーの程度を示すために使用される. ガウシアンフィルタに基づく緑から赤への連続的なカラーグラデーションにより、エラーなしから最大閾値までの不一致の程度を表す.

フレーム内差分 手首の可視化において、特定のフレームで異なる部分を識別するため、異なる色の骨と関節が使用されている (図 1 右下). 不一致の度合いに応じて、関節や骨格を含む異なる部分の色は緑から赤へと徐々に変化する.

2.3.2 視線入力

提案システムでは、視線による操作で、再生速度と表示される手を変更することができる (図 1 右上). ユーザの視界の上部には 4 つの黄色い立方体があり、それぞれ異なる値 (0.25, 0.50, 0.75, 1.00) がラベル付けされている. ユーザが立方体を注視すると、徐々にロードされる青い円が表示される. 1 秒後に円が完全に埋まると、黄色の立方体が青色に変わり、対応する機能が起動する. これにより再生速度が選択された値に変更される.

同様に、右側には異なる単語 (both, teacher, student) でラベル付けされた 3 つの立方体があり、それぞれお手本と生徒の手の両方、お手本の手のみ、生徒の手のみを表示する設定が可能である.

3 評価実験

12 名のピアニスト (女性 9 名, 男性 3 名, 年齢 20 ~ 28 歳, 平均 23.9, 標準偏差 2.6) が本実験に参加した. 参加者のうち、全員が 13 ~ 23 年 (平均 18.8, 標準偏差 2.6) のピアノ学習経験があった.

3 つの訓練条件による被験者内デザインを用いた: (1) 録画したビデオ授業のみ (Video), (2) タッチスクリーンによる視覚化 (Screen), (3) 提案システム (AR).

参加者は、上記の条件をシャッフルした順序で体験し、トレーニングセッション前後のパフォーマンスを記録して定量的評価を受けた. 定性的評価には NASA-TLX [6], System Usability Scale (SUS) [7],

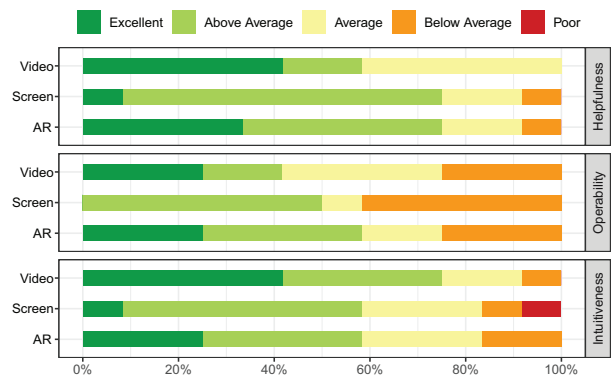


図 2. アンケートに対する参加者の回答. 3 つの条件の有用性, 直感性, 操作性についての 5 段階リッカート尺度による回答結果.

User Experience Questionnaire (UEQ-S) [8] 質問票を用いた.

4 結果

Screen と AR のユーザビリティを測定するために、SUS [7] を使用した. Screen の結果が 64.59 であったのに対し、AR の結果は 68.96 であり、平均 (68.00) 以上の許容範囲であった. UEQ-S [8] の結果は、実用品質、快楽的品質、総合得点に関して評価する. 実用品質では Video が最も高いスコア (平均 1.396) を得た. 快楽的品質 (平均 2.479) と総合得点 (平均 1.823) については、AR が最も優れていた.

アンケートに対する参加者の回答を図 2 にまとめた. 有用性、直感性、操作性の観点では、Screen よりも AR が優れていた. 以下のような理由がある. AR の環境での 3D の可視化は、より直感的で視覚的に魅力的であったためである. AR の速度変更や表示する手を変更するインタラクションは理解しやすく、同じ時間内に操作を習得できる. AR による没入的学習環境がユーザの学習とシステムへの理解を容易にしたと示唆される.

5 まとめ

本論文では、ピアノ学習のための新しい拡張現実システム「PianoSyncAR」を提案し、AR における不一致の可視化がどのように設計され、評価されたかのワークフローを示した. 結果分析からの肯定的なフィードバックに基づき、提案システムは、既存のビデオベースの方法と組み合わせてピアノ学習を強化するための貴重な補完ツールであることが確認された. さらに、AR で表示される不一致の視覚化手法は、技能習得の他の分野にも有用な洞察を提供できると考えられる.

謝辞

本研究は JST CREST JPMJCR17A3 および JST ムーンショット型研究開発事業 JPMJMS2012 の支援を受けている。

参考文献

- [1] Mitsuo Kawato and Hiroaki Gomi. The cerebellum and vor/okr learning models. *Trends in Neurosciences*, 15(11):445–453, 1992.
- [2] Shinichi Furuya. Individual differences in sensorimotor skills among musicians. *Current Opinion in Behavioral Sciences*, 20:61–66, 2018. Habits and Skills.
- [3] Takanori Oku and Shinichi Furuya. Noncontact and high-precision sensing system for piano keys identified fingerprints of virtuosity. *Sensors*, 22(13), 2022.
- [4] Debidatta Dwibedi, Yusuf Aytar, Jonathan Tompson, Pierre Sermanet, and Andrew Zisserman. Temporal cycle-consistency learning. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [5] Donald J. Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining, AAAIWS'94*, page 359–370. AAAI Press, 1994.
- [6] S. G. Hart. Nasa task load index (tlx). volume 1.0; paper and pencil package. 1986.
- [7] John Brooke. Sus: A quick and dirty usability scale. *Usability Eval. Ind.*, 189, 11 1995.
- [8] Andreas Hinderks, Martin Schrepp, and Jörg Thomaschewski. A benchmark for the short version of the user experience questionnaire. 09 2018.

未来ビジョン

「PianoSyncAR」システムはその有効性を実証したが、さらなる改善の機会を認識することが重要である。しかし、拡張コンテンツを提示するためには、重く嵩張るヘッドマウントディスプレイ (HMD) を用いることが必要である。重くてかさばるヘッドマウントディスプレイ (HMD) に依存していること。参加者の一人は、HMD を長時間装着することに伴う不快感や疲労について懸念を表明しており、これは長時間の拘束を伴うピアノの練習セッションでは特に問題となりうる。この問題に対処するため、拡張現実感機能を備えたスマートグラスやコンタクトレンズなど、ウェアラブル技術の今後の進歩により、より快適で軽量の代替手段を

提供できる可能性がある。このような進化を取り入れることで、「PianoSyncAR」システムの使いやすさとユーザーエクスペリエンスがさらに向上し、ピアニストにとってより実用的なものになると考えている。

最終的な目標は、あまり専門的でないユーザーでも、家庭で活用できる MR ピアノトレーニングシステムを開発することである。さらに、動作の同期、不一致の可視化、MR 世界での表示といったプロセスを、技能習得の他の領域にも拡張することができると思う。