第 28 回日本バーチャルリアリティ学会大会論文集（2023 年 9 月）

Analyzing VR Game Replays to Uncover Skill Progressions and Enable Personalized Skill Transfers

Zhixing Wang[1)], Keigo Matsumoto[1)], Hideaki Kuzuoka[1)], Takuji Narumi[1,2)]
1) 東京大学大学院情報理工学系研究科（〒 113-8656 東京都文京区本郷 7-3-1,
{zhixing, matsumoto, kuzuoka, narumi}@cyber.t.u-tokyo.ac.jp）2) JST さきがけ

Abstract: Popular VR games such as Beat Saber gathered terabytes of game replays, with clear motor skills shown by hand speeds and head angles. To unlock more insights into VR skill acquisition, we cluster the motion by latent vectors from an auto-encoder model. We identify a diverse progression of motion motif usage from novices to experts. We also show the model's potential in predicting players' accuracy improvements, hinting at new model-based skill transfer methods.

Keywords: Motor learning, Auto-encoders, Skill progression, VR Games

1. Introduction

From a toddler's clumsy steps to a golf player's confident swing, learning a motor skill is not always easy. Lucky for us, we can also help ourselves with the potential of virtual reality (VR) and machine learning (ML) technology. By analyzing and manipulating the interaction between users and the virtual environment, we could better understand and help motor skill learning in VR.

Existing research for skill transfer in VR seems to center around motion recordings from experts or replays. As demonstrated in the CanKendama study [1], recorded expert motion, in addition to slower speed in the virtual environment, helped novices learn a not-so-simple Kendama trick. In more professional settings, SPinPong [2] successfully applied visual, haptics, and temporal guidance to returning spin shots in table tennis. This study found that visually displaying a recorded racket over the player's body works the best among tested conditions.

Going beyond recorded motion, a recent study [3] also explored active skill transfer methods using "virtual co-embodiment". In the study, teachers and students shared control over an avatar, performing a demanding dual task where each hand is reaching a different target simultaneously. Here, the co-embodiment scheme outperformed overlayed movements, suggesting the need beyond simply displaying the teacher's movements when the task is complex enough to require integrated, unconscious moves.

In addition to in-game applications, game replays could also provide insights on skill learning. For example, Listman et al. [4] examined performance data on Aim Lab[1], an FPS (First Person Shooting games) trainer. The results revealed an imbalance between acuity and accuracy improvements, which concurred with existing literature and validated the use of gameplay data for motor learning studies. A more recent example looked at Beat Leader[2], a community leaderboard for the VR rhythm game Beat Saber. In the game, players cut approaching blocks to the beats with their lightsabers, generating motion data as they play. From the data, the study[5] accurately identified over 50,000 players using only their head and hand movements. Besides biometrics data such as height and arm length, movement patterns are also key to their model, suggesting a personalized skillset.

Following these works, we analyze the same Beat Leader replay dataset further to understand how experts play the game and how novices could learn to get there. We first look for simple metrics to validate and argue for the existence of highly skilled experts in the game (RQ1). Then, we use an auto-encoder model to reveal how the patterns of motions change along ranks, hoping to understand how one might become a master (RQ2). Finally, we look at the increment steps along this path towards mastery, and ask how well our model identifies actions that lead to accuracy gain(RQ3).

Key findings from our analysis are:

- Experts have higher hand movement speeds and better head synchronization with incoming blocks.
- Novices start from similar clusters of movements, but experts become more diverse in their profiles.
- Encodings from the same player predict accuracy gain better than baseline and cross-player ones.

We then discuss how the results could lead to more effective and personalized skill transfer methods in VR.

[1]https://www.aimlab.gg/

[2]https://www.beatleader.xyz/

Table 1: Dataset Statistics

Subset	No. Scores	Average Accuracy (± std)
Full	4,227,434	80 % (± 17)
1	652	74 % (± 14)
2	5,763	81 % (± 15) before 85 % (± 12) after

2. Methods

2.1 Dataset

The official Beat Saber game doesn't support leaderboards on custom songs, so enthusiasts come up with their own systems. Beat Leader is a community leaderboard launched in early 2022. It aims to be transparent, open-sourcing all components and importantly features a high-quality replay module. The openly available replays make it well-suited for our purposes. We obtained 4,227,434 scores (gameplays) from the public API in 2023-04 (also dumped here[3]). Each score is linked with a leaderboard/map, a player, and a replay.

Leaderboards are centered around a song map on which all scores from the board are obtained. The 'map' is a game level that includes a song track, a set of block positions syncing to the beats, and other additional gaming elements such as walls, bombs, or other extensions.

Players cut approaching blocks to get scores based on their timing, positional accuracy, and magnitudes of swings. Besides the accuracy of a single cut, 'accuracy' is also used to describe the overall score on a song, defined as a player's score divided by the maximum possible.

Finally, replays, as introduced before, are the recorded motion data of the player throughout the map, containing positions and rotations for the head and hands. The entire dataset consists of terabytes of recordings, reaching far beyond our computational budgets at the time. Therefore, we have decided to narrow down to the following two subsets of data: first on a single, well-designed map to understand the expert's motion, then a set of successful improvements across all maps to study the increment of accuracy. Counts and score accuracy statistics are shown in Table 1.

2.1.1 Subset 1: One particular map

To help people find suitable maps to play, the Beat Leader team maintains a rank system for map difficulty. Each popular map the community suggests will go through an objective ranking process[4] to determine its playability and difficulty. To ensure a good basis for analysis, we start from such a ranked map.

The rank of a map is represented in stars. From the easiest (0-1 stars) to the craziest maps (14 stars), all players should be able to find a good quality map at their level. For our purpose, we choose to select a moderately challenging map at 8.3 stars total[5]. As shown in Table 1, its score accuracy is lower than other subsets, indicating its difficulty and hence suitability for expert analysis.

2.1.2 Subset 2: Improvements over all maps

To find out a player's improvements over time, we need to have multiple recordings of the same player passing the same song map. However, due to collisions of replay names (which only contain playerId and leaderboardId), newer replays always override previous ones. Fortunately, over a short period of API change (from 11 Mar 2023 to 11 Apr 2023), BeatLeader briefly included scoreId in replay names, leaving a full history of 244,759 records.

Out of the 245k fully named replays, only 5,763 are repeating entries coming from the same player on a map, and only 31 of them are three-time challengers. Note this is a very small subset of scores (only about 1% during the period), suggesting either the scarcity of super enthusiastic players or the difficulty of song maps in general.

We group the entries by player and map and get 2,866 records of improvements. We use the first and last scores within a group as the performance before and after improvement, shown in Table 1. While the average improvement is only 4%, considering the starting point, this is a hard-won achievement. For this report, we further filter through the subset, and focus only on the improvements that have more than 10% of accuracy gain, to reveal how a novice managed to vastly improve their skills.

2.2 Analysis & Model

For RQ1, we look at basic metrics about a player's movements. In particular, we calculate the average speed for each hand to study motor execution; we also compute the correlation between head direction and incoming block position to measure input responses. We then compare the metrics between an expert group (top 10 players in Subset 1) and a novice group (bottom 10), thus giving basic characteristics about an expert's move.

To further understand the complex motion patterns learned by these master players, we seek help from powerful ML and deep-learning techniques. Previous works [6, 7] had successfully applied neural networks to high-level identification of motion data. In these studies, the model could represent a short window of movements in succinct codes or a latent vector. Clustering on such vectors gives labels of recurring units of movements, or motion motifs. The distribution of motion motif used in a clip, then, would become the signature of it. For example, it can indicate the type of dance performed, or the

[3] https://discord.gg/2RG5YVqtG6
[4] https://beatleader.wiki/en/criteria
[5] https://www.beatleader.xyz/leaderboard/global/4a2a91/

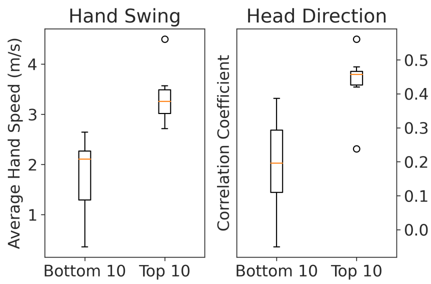

Fig. 1: Boxplots of movement metrics

behavior of an explorative mouse.

For this study, we adapted the auto-encoder model in [7], which includes both an encoder and a decoder. This is preferable for our application since the decoder can generate novel motion sequences to suit each player.

Applying the model to RQ2, we calculate latent vectors and motif labels for each frame of the replays in Subset 1. Since the labels don't necessarily correspond to well-defined moves without properly tuning the algorithm, we focus on the overall distribution as the signature of each playthrough, as in [6]. Then, we use entropy to measure how concentrated the distribution is, and group by player's rankings on scores, thus describing how experts utilize a variety of motifs in contrast to novices.

For RQ3, we compare distance measures for accuracy improvement prediction (by R2 values on linear regressions), so one can make changes guided by such measures. We first compute a baseline: Euclidean distance on position and rotation vectors[6]. We then compute distances on encoded latent vectors, both for same-player (Subset 2) and for cross-player accuracy differences. For the cross-player setting, we resample Subset 1 to generate a set of dummy improvement pairs whose distribution of accuracy gains matches Subset 2 but are from two players, to explore whether player identity affects the metric.

3. Results

3.1 RQ1: Indication of fine motor skills

As Figure 1 shows, the top 10 expert players have a clear lead in average hand speed and head direction correlation with incoming blocks. This shows better motor execution befitting the game's fast pace and also a more integrated execution indicated by head movements coordinated with the next target to cut. Novices also seem to have larger variance in both movement metrics, but does this mean they have more variance in "skills"? We move on to the latent space to find out.

3.2 RQ2: The road towards mastery

In Figure 2, we plot the motif usage of 31 groups of players from the lowest ranks (left) to the highest (right) as a heatmap. Each column of the heatmap is the average distribution for one group of players (21 players per group, arbitrarily chosen for presentation). The color at each cell along this vertical strip (15 motifs in total, default parameter of [7]) indicates the proportion of frames clustered under that particular motif. For example, the lowest ranked groups concentrate on clusters 13 and 14, using each of them about 25% of the time, whereas highly ranked groups seem to have a more dispersed span over all 15 motifs, only moderately favoring one or two.

Looking at the figure, instead of a ladder - a progression of successive motion units as one moves up the rank of scores - we see a landscape, a diffusion-like process where players start from a rather limited set of "novice" motions but then spread out across the entire space of motifs as they advance in ranks. This is also quantitatively measured by the entropy for each group, with a strong linear fit between entropy and rank (Figure 3, slope=0.013/group) - an increasing diversity trend.

Our results here still look at experts and novices as groups. We now know that experts are more diverse in their skill sets, but how did each of them get there? Can our model help understand the incremental steps in an individual player's journey toward mastery?

3.3 RQ3: Predictive metrics in the model

In this section, we look at the predictive power of our model's latent codes on accuracy improvements. Note that we didn't train the model to predict accuracy or any goal-related information: its only target is to reconstruct and predict the motion itself. So, any weakly predictable result can be seen as indicative of future development, should we include such additional targets.

As listed in Table 2, out of the three distance met-

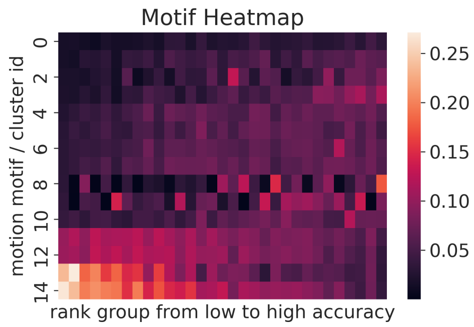

Fig. 2: Heatmap of motif distribution by rank

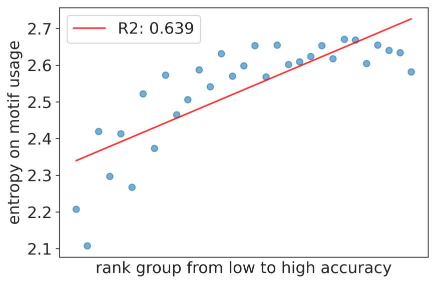

Fig. 3: Best line fit for motif entropy by rank

[6]as saber-tip positions to avoid erratic change at 180 °

Table 2: R2 Values predicting Accuracy Gain

Measure	R2 Value
Baseline (Same-player)	0.00
Latent (Same-player)	**0.11**
Latent (Cross-player)	0.03

rics we tested, latent codes on the same player is the only one that can weakly predict the accuracy gain after an improvement. This suggests its potential application for learning, where learners would identify movements to improve by looking at where the latent distance is the largest between their recordings and the teacher.

On the other hand, the baseline metric is essentially useless when predicting improvements. So, if we base a guidance method on the raw distance between the learner and the teacher, it won't be much effective since closing large gaps in head and hand positions doesn't necessarily lead to increased performance, at least in BeatSaber.

Finally, while the cross-player metric is better than the baseline, it's much worse than same-player pairs. This further validates our understanding of the progression of skill improvements in RQ2: When following a player's own development path, the further you go, the more you improve; on the contrary, the distance between two arbitrary replays by two players will mix up their paths and thus tells little about their accuracies.

4. Discussion

Our results first validate the use of non-controlled replay data to extract characteristics of motor skills in a VR game, from the speed of hand swings to synchronized head movements with incoming action targets. Compared to non-VR video games, typical VR devices can record more sensor data at high framerates, allowing for more detailed analysis. We hope this will draw more attention to such open data and consequently produce more insights to help motor learning and beyond.

We then focus our analysis on the learning path of a player's movement patterns. We use an existing autoencoder model to generate cluster labels and study their distributions. One possible limitation is the lack of tuning and adaptation of this model, causing some players only to have one single motif assigned to their entire replay, which is highly unlikely considering the complexity of their movements. Nevertheless, the group-wide statistics remain useful for an initial understanding of the progression of motor motif usage.

Across this diverse landscape of skill progression paths, experts don't all flock towards one single peak of mastery - they march forward on an individual trek of self-improvement and personal journey. It's a more dynamic view of the learning process and hopefully will lead to more guidance designs that focus on individuality instead of a single template of expertise.

In the end, the preliminary results on our model's predictive power are only a hint at how it could be used for such new methods. For example, one can use a model-based distance to change control weights in a co-embodiment setting dynamically. Or even better, we could deploy the generative model to override our movements with a personal motion output, as if we **ourselves** were one rank higher than our current level. Imagine embodying a future expert version of yourself - it would be a liberating learning experience.

Acknowledgement The authors would thank Chi-Lan Yang and Reigo Ban for their helpful ideas and feedback. We also thank Viktor Radulov (@NSGolova) and others at BeatLeader for building the leaderboard and kindly answering our questions. This work was partially supported by Grant-in-Aid for Scientific Research (S) (19H05661), JST Moonshot Research & Development Program (JPMJMS2013), and JST ACT-X Grant Number (JPMJAX22AG).

References

[1] 川崎仁史, 脇坂崇平, 笠原俊一, 齊藤寛人, 原口純也, 登嶋健太, 稲見昌彦, *et al.*, "けん玉できた! vr: 5 分間程度の vr トレーニングによってけん玉の技の習得を支援するシステム," **エンタテインメントコンピューティングシンポジウム *2020* 論文集**, vol. 2020, pp. 26–32, 2020.

[2] E. Wu, M. Piekenbrock, T. Nakumura, and H. Koike, "Spinpong-virtual reality table tennis skill acquisition using visual, haptic and temporal cues," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 5, pp. 2566–2576, 2021.

[3] D. Kodama, T. Mizuho, Y. Hatada, T. Narumi, and M. Hirose, "Effects of collaborative training using virtual co-embodiment on motor skill learning," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, no. 5, pp. 2304–2314, 2023.

[4] J. B. Listman, J. S. Tsay, H. E. Kim, W. E. Mackey, and D. J. Heeger, "Long-term motor learning in the "wild" with high volume video game data," *Frontiers in human neuroscience*, vol. 15, p. 777779, 2021.

[5] V. Nair, W. Guo, J. Mattern, R. Wang, J. F. O'Brien, L. Rosenberg, and D. Song, "Unique identification of 50,000+ virtual reality users from head & hand motion data," *arXiv preprint arXiv:2302.08927*, 2023.

[6] A. Aristidou, D. Cohen-Or, J. K. Hodgins, Y. Chrysanthou, and A. Shamir, "Deep motifs and motion signatures," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 6, pp. 1–13, 2018.

[7] K. Luxem, P. Mocellin, F. Fuhrmann, J. Kürsch, S. R. Miller, J. J. Palop, S. Remy, and P. Bauer, "Identifying behavioral structure from deep variational embeddings of animal motion," *Communications Biology*, vol. 5, no. 1, p. 1267, 2022.