Do motion retargeting and AI-generated motion preserve anticipatory visual information in boxing animation?

Annabelle Limballe
M2S-EA 7470
Univ Rennes, Inria
Rennes, France
annabelle.limballe@univ-rennes2.fr

Ahmed Abdourahman Mahamoud
IRISA-UMR 6074
Inria, Univ Rennes, CNRS
Rennes, France
ahmed.abdourahman-mahamoud@inria.fr

Guillaume Claude
M2S-EA 7470
Univ Rennes, Inria
Rennes, France
guillaume.claude@univ-rennes2.fr

Simon J. Bennett
Research Institute for Sport and Exercise Sciences
Liverpool John Moores University
Liverpool, United Kingdom
S.J.Bennett@ljmu.ac.uk

Richard Kulpa
M2S-EA 7470
Univ Rennes, Inria
Rennes, France
richard.kulpa@univ-rennes2.fr

Franck Multon
IRISA-UMR 6074
Inria, Univ Rennes, CNRS
Rennes, France
franck.multon@inria.fr

Abstract

In combat sports simulation, expert users have extensive experience in extracting visual anticipatory information from their opponent's movements, which should be preserved in virtual opponent animations. However, animations based on motion capture data rely on complex pipelines that may disrupt the visual information initially present in the real actor's motion. Recent physics-based simulation (such as Adversarial Motion Prior, AMP) have recently been introduced to adapt character motion to interactive constraints while satisfying physical laws, but they may introduce additional distortions. In this paper, we examine how animating a virtual opponent using motion capture data (as a reference), retargeting techniques, and AMP influences the visual anticipatory performance of expert boxers. Ten boxers participated in a perceptual study using video clips with varying occlusion times, in which they were asked to predict the type of punch among four possible choices. The results show a clear decline in anticipation perceptual accuracy (PA) as occlusion occurs earlier. Hooks were more difficult to anticipate than straight punches (PA for rear hook = 0.455, lead hook = 0.435, rear straight = 0.660, lead straight = 0.744). Animations based on motion capture (mean accuracy = 0.758) led to significant better anticipation perceptual accuracy, for straight punches, compared to retargeting (mean accuracy = 0.700) and AMP (mean accuracy = 0.647). Overall, these results highlight the need to redesign traditional animation pipelines to better preserve the visual information picked up by expert users in virtual training simulators.

SAP '26, Rennes, France

ACM ISBN 979-8-4007-2799-3/26/08
https://doi.org/10.1145/3821409.3821464

CCS Concepts

• **Software and its engineering → Virtual worlds training simulations**; • **Computing methodologies → Motion processing**; **Physical simulation**; **Perception**.

Keywords

Motion simulation, imitation-based animation, motion retargeting, visual information, anticipation skills

ACM Reference Format:
Annabelle Limballe, Ahmed Abdourahman Mahamoud, Guillaume Claude, Simon J. Bennett, Richard Kulpa, and Franck Multon. 2026. Do motion retargeting and AI-generated motion preserve anticipatory visual information in boxing animation?. In *ACM Symposium on Applied Perception 2026 (SAP '26), August 26–27, 2026, Rennes, France.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3821409.3821464

1 Introduction

Boxers must be able to adapt quickly to changing events: locate information in their visual field, anticipate their opponent's actions, and react to external stimuli. Perceptual abilities are therefore essential to boxing performance [Hausegger et al. 2019; Mann et al. 2007; Romeas et al. 2022]. These perceptual skills are difficult to analyze and train. Virtual reality (VR) offers an opportunity to analyze perceptual skills in sports [Bideau et al. 2010; Michalski et al. 2019], in standardised situations, while aiming at preserving the natural perception-action coupling [Gibson 1979]. However, in VR, the opponent is simulated using virtual human, with different appearance, morphology and skeleton. Hence, in the context of serious VR fighting sport training, a specific attention should be paid to preserve the information naturally used by the users, to maximize the transfer of skills to real-world practice [Michalski et al. 2019; Pinder et al. 2011]. The animation pipeline generally relies on motion capture data, which are very difficult to get in the specific case of fighting sports. Most of the motion capture systems involve placing sensors or markers on the boxer's body that may disrupt the way they naturally perform their motion, and generate experimental challenges. These systems track a full body model of the subject based on sparse marker or sensor data, using various model-fitting methods, introducing potential inaccuracies. In the VR training simulator, the animated virtual character may have different skeleton structure, dimension, and geometric envelope, which involves adapting the captured motions to these new constraints, using rigging and motion retargeting. These methods rely on constraints solving, with heuristics, such as minimizing joint angle changes, minimizing foot sliding, etc. However, nothing ensures that the resulting motion preserves the visual information (VI) required for users' natural perceptual strategies. Moreover, the virtual opponent does not simply replay open loop animations, but should react to the users actions. To address this problem, recent works [Peng et al. 2021; Younes et al. 2023] have proposed to use physics-based imitation learning. These methods aims at imitating a small set of motion trials, which have been retargeted to a unique physics-based skeleton. The joint forces are computed using a Proportional-Derivative controller (PD-controller) driven by key poses computed at each frame. The aim of the imitation-based controller is to compute the relevant key poses to generate motions that fall into the statistical distribution of the training samples. However, PD-controllers can generate biomechanically unfeasible joint torques [Xie et al. 2023], although the resulting motion globally imitate the motion capture data. The VI used by experts might consequently be disrupted with this method.

In this paper, we aim at evaluating if these state of the art character animation methods preserve VI, especially in fighting sports.

2 Related work

2.1 Anticipatory visual information

From a sports science perspective, to satisfy the physical, temporal, psychological, and event constraints [Michalski et al. 2019; Pinder et al. 2011], it is essential to preserve the kinematic VI in the virtual environment, used for anticipation. In fighting sport, the time constraints are so tight that a boxer must take early VI allowing the anticipation of the upcoming opponent's action. Anticipation is defined as a successful prediction of the future action before it has ended or even started, based on kinematic and contextual information [Müller et al. 2022]. Experts take the more appropriate information, earlier in the action, compared to the novices [Williams and Jackson 2019]. Most of this information is given by VI that boxers has to pick-up to anticipate the opponent's action.

A traditional method to investigate VI is video occlusion [Song et al. 2025; Williams and Jackson 2019]. A video clip of an opponent performing a key action is presented to participants with various temporal or spatial occlusions, and the participant has to guess the outcome of the action. The accuracy of the participant's response is then used to determine the moment and/or the body part bearing the essential VI supporting anticipation.

In a VR boxing simulator, maximizing skill transfer from virtual to real-world practice requires accurately preserving VI to avoid disrupting the natural perception–action loop. Hence, the occlusion paradigm should yield similar results whether using real video footage or simulated avatars in VR. Some studies have experimentally disrupted movements in VR (motion dynamics, avatar or objects size, muscle contraction, etc.) [Kenny et al. 2019; Reitsma and Pollard 2003; Yamac et al. 2024]. These studies highlighted participants' sensitivity to variations in motion kinematics. In our study, we use this sensitivity to assess whether kinematic differences are perceived between different animation methods.

2.2 Character animation pipeline

Motion capture joint angles are consistent with a given character model (skeleton, dimension, and 3D envelope) designed for the actor. Directly applying such joint angles to a new character generally leads to artifacts, such as foot skating, self-penetration, geometric constraints loss…Motion retargeting aims at solving a set of manually designed constraints, such as preserving contacts [Gleicher 1998; Kulpa et al. 2005]. Interaction meshes propose to automate the design of these constraints [Ho et al. 2010], with an extension to surface points [Basset et al. 2020; Jin et al. 2018]. However, all these optimization-based methods do not guarantee to preserve the initial features of the captured motion. Recent studies have proposed the use of recurrent neural networks [Villegas et al. 2018] and autoencoders [Aberman et al. 2020; Hu et al. 2023], trained on large datasets of movements applied to various body types, with the aim of capturing these intrinsic characteristics. This approach has also been extended to deal with surface-to-surface contact points [Cheynel et al. 2025; Jang et al. 2024; Ye et al. 2024]. Whatever the approach, the main constraints focused on contact preservation, and fixing geometric artifacts, but there is no explicit constraints to preserve VI.

To adapt the retargeted motion to new situation, imitation learning has become a very popular approach. It consists in imitating motion capture data while taking new kinematic and dynamic constraints into account. These methods generally use PD-controllers to generate the joint torques required to imitate a motion, but with potentially unrealistic forces [Xie et al. 2023]. Imitation learning is based on minimizing the error between the simulated poses and example poses [Lee et al. 2010; Liu et al. 2016; Peng et al. 2018], but with limited ability to adapt to various conditions. Adversarial imitation learning [Ho and Ermon 2016; Xu and Karamouzas 2021] is an alternative approach to avoid manually designing and tuning specific pose error metrics. This approach relies on an adversarial discriminator, aiming at distinguishing simulated motions from those given in the demonstration dataset. Adversarial Motion Priors (AMP) [Peng et al. 2021] and its improvements [Bae et al. 2025; Luo et al. 2024, 2023; Pan et al. 2025; Peng et al. 2022; Tessler et al. 2023; Xu et al. 2025; Younes et al. 2023] proposed a stable and efficient solution to more accurately control the character. However, the reward used for training generally relies on a global metrics for sequences composed of multi-dimensional state vectors. It could lead to a good reward even though a few body parts may exhibit important errors, and subtle velocity profiles may be badly simulated. Moreover, the demonstrations used during training are based on retargeted motion capture data, potentially introducing errors. This accumulation of inaccuracies may disrupt relevant VI in anticipation tasks.

3 Methods

In this paper, we investigate whether state-of-the-art character animation methods disrupt the VI that experts naturally use to

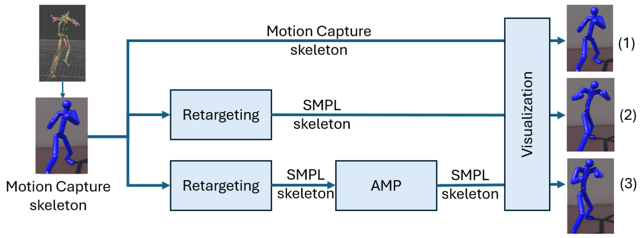

Figure 1: Overview of the experiment. Motion capture skeleton at T1, 1) directly visualized by the participants, 2) retargeted to a SMPL skeleton, and 3) used to train an AMP dedicated model with SMPL-compliant skeleton. The resulting poses slightly vary depending on the animation method.

anticipate an opponent's actions in fighting sports. To this end, we adopted a video occlusion paradigm commonly used in psychophysics. Four types of punches were captured from an expert boxer and processed using three different animation methods (see Figure 1): 1) tracking a motion-capture skeleton using a commercial system (Qualisys QTM), 2) retargeting the motion-capture data to an SMPL-compliant skeleton using a standard retargeting method [Luo et al. 2023], and 3) training an AMP model [Luo et al. 2023] to imitate the retargeted SMPL-compliant motion.

3.1 Virtual character models

Motion capture data of an expert boxer (left-handed) was collected using the Qualisys optical motion capture system. 43 reflective markers were placed over standardized anatomical landmarks ("Sports Marker-set" Qualisys model). For the reconstruction and cleaning processes, we used the standard Qualisys pipeline, including the definition of an AIM rigid-body model and polynomial interpolation for badly reconstructed markers (for maximum 10 incorrect frames at 100Hz). No trajectory filtering was applied. The Qualisys processing pipeline was also used to track a skeleton (denoted the motion capture skeleton) composed of 24 joints with a total of 72 degrees of freedom (3 for each joints, see Figure 6). The boxer was asked to perform 4 types of punch (lead/rear straight, lead/rear hook, rear hook) aiming at a pad held by an instructor. Each motion type was repeated 10 times. Preliminary to the punch, the boxer was asked to perform the same idle motion, in all the trials, to avoid giving information before the last displacement involving the punch (see the accompanied video). Two animation methods, namely motion retargeting and AMP [Peng et al. 2021], were used to animate a SMPL model [Luo et al. 2023] (see Figure 6), with a similar number of joints (24 joints/72 DoF). Choosing a very different skeleton (too much simplified, or too much complex) would lead to unfair comparison with the motion capture skeleton.

To ensure a fair comparison between the three conditions (motion capture, retargeting and AMP), we did not used any supplementary retargeting or post-processing. The motion capture and SMPL skeletons were rendered using a simple stickman representation: each joint is rendered by a simple sphere, and the parent–child relationship between two joints was represented by a simple 3D capsule, whose length is defined by the relative position of the child joint with respect to its parent.

3.2 Retargeting

In this paper, retargeting is used to adapt the poses of the motion capture skeleton to the SMPL one. We used the retargeting method commonly used in imitation physics-based simulation [Peng et al. 2022, 2021; Tessler et al. 2023; Younes et al. 2023]. It is composed of 5 steps, in a pose-by-pose manner: 1) Definition of a joint correspondence dictionary between the Qualisys and the SMPL skeleton (see Table 1); 2) the motion capture pose and its corresponding T-pose are aligned with the SMPL skeleton (also in a T-pose); 3) the root translation is adapted to the size of the SMPL skeleton; 4) and the global rotation of all the joints (given relatively to the motion capture skeleton T-pose) are directly applied to the SMPL skeleton.

3.3 Imitation-based virtual human simulation

In AMP [Peng et al. 2021], imitating the expert motions/demonstrations becomes a reinforcement learning problem. The original method was based on a specific and simplified virtual human model, and further adapted to deal with the SMPL skeleton [Luo et al. 2023]. Given an observation of the character's current state s_t, the policy π_θ generates an action a_t that changes the character's state to maximise the similarity reward r_t between the simulated motions and the expert motion capture demonstrations M_E: $s_{t+1} = \pi_\theta(a_t|s_t)$. In this model, a_t corresponds to target joint angles (target poses) applied to a Proportional Derivative (PD) controller aiming at computing the corresponding joint torques, for each joint. The gains of the PD controllers were set as specified in [Peng et al. 2021]. The state s_t describes the configuration of the character's body: the relative positions of each joint with respect to the root (pelvis), their rotations (quaternions), and their linear and angular velocities.

In an adversarial imitation context, a discriminator is trained to generate the reward r_t that guides the policy to imitate the expert demonstrations from the motion M_E. As presented in [Peng et al. 2021], the policy and the discriminator were made of a single multilayer perceptron with two hidden layers of sizes 1024 and 512. As we wish to evaluate AMP in its most favourable condition, we trained a separate policy to imitate each studied motion.

3.4 Experimental set-up

In this study, we aimed to evaluate if motion retargeting and physics-based imitation models preserve anticipatory VI, using the occlusion paradigm: assess if expert participants have different anticipatory performance depending on the animation method.

10 participants (2 female, 8 men, age 30.6 ± 7.6 years) completed the experiment. They were boxers practicing French boxing, with 9,7±7 years of competitive sports experience. They performed at various levels of competitions (from regional to international). All participants were healthy, without injuries and had normal or corrected-to-normal vision.

3.4.1 Materials and task. The participants were seated in front of a computer screen, 30 cm away from the screen (27"/4K/144 Hz). To ensure precise and equivalent display for each motions, videos were rendered exactly at each frame of the motion, matching the 100Hz motion capture data. The experiment was built on PsychoPy software [Peirce et al. 2019]. The participants had to predict which type of punch was presented in the video, among four possibilities:

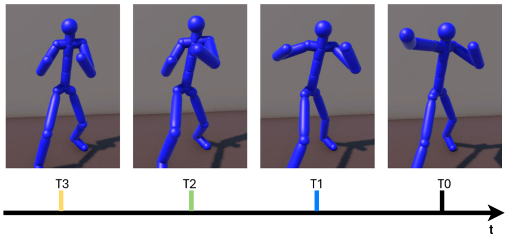

Figure 2: Decomposition of the motion clips into sub-clips separated by events in the punch sequence: T_3 (last leg extension before the punch initiation), T_2 (punching arm movement initiation), T_1 (2/3 of the time after punching arm movement initiation) and T_0 (impact), for a lead straight.

lead straight, rear straight, lead hook, and rear hook. The response was recorded using the 1, 2, 3 and 4 keys on the computer keyboard. Participants had 5s to answer, to simulate timing pressure.

The motion capture data were processed using the AMP pipeline: retargeting to the SMPL skeleton, and imitation reinforcement learning to design a physics-based controller, as described above. Finally, 144 video clips were designed for the computer graphics occlusion experiment: 3 animation methods (motion capture, retargeting, AMP) x 4 types of punch (lead straight, rear straight, lead hook, rear hook) x 3 occlusion times (leading to three animation intervals $[T_s - T_1]$, $[T_s - T_2]$, $[T_s - T_3]$) x 4 repetitions of each type of punch. This block of 144 trials was designed with a randomized order. Another similar block, with different randomized order, was designed to propose two sessions of 144 trials separated by a break in-between, leading to a total of 288 trials.

The 3 occlusion times were determined relatively to the punch impact T_0. T_3 was designed as the last support leg's extension preceding the punch (forward movement of the punching hand). T_2 was designed as the moment the punching hand started moving (exited the guard position). T_1 was set at 2/3 of the time between T_2 and the punch impact T_0, delivering information about the beginning of the punching hand trajectory. T_s was the start of the clip corresponding to the guard phase preceding the punch. During the experiment, the participants could consequently see short starting sequence $[T_s - T_3]$, up to longer $[T_s - T_1]$ timing interval. Figure 2 depicts the different occlusion times, for a direct punch. Before the experimentation, a familiarisation phase was proposed, presenting complete motion capture sequences, from T_s to T_0. The total experiment, including the familiarisation phase took 40 min.

3.4.2 Data analysis. All statistical analyses were performed in R Studio (version 2026.01.2). Participants' responses were coded as a binary response for each trial (1 = correct, 0 = incorrect) and used to compute the perceptual accuracy defined as the number of correct responses among the total responses. It will be denoted as $PA_{type\ of\ punch,\ animation\ method}(occlusion\ time)$ throughout the rest of the paper. We examined whether perceptual accuracy in each condition exceeded the theoretical chance level, defined as 25% (four possible choices). For each combination, we used an unilateral t-test (H_1: accuracy > chance level), with Bonferroni correction for multiple comparisons. To take into account the effect size, we computed the Cohen's d value for each condition as the difference between the observed mean accuracy and the chance level (25%), divided by the observed standard deviation.

Generalized Linear Mixed Models (GLMMs) were used to investigate differences between conditions, using the lme4 package, specifying a binomial distribution with a logit link function. The occlusion time, the type of punch and the animation method were entered as fixed effects. A by-participant random intercept was included to account for individual variability in baseline response accuracy. Post-hoc pairwise comparisons were conducted using the emmeans package. Rather than exhaustively testing all possible pairwise combinations, comparisons were restricted to theoretically meaningful contrasts, based on the experimental hypotheses. For the decomposition of the three-way interaction, only comparisons between the animation methods within each combination of occlusion time and type of punch were performed. Bonferroni correction was applied across this restricted set of comparisons. For main effects, all pairwise comparisons between levels were conducted with Bonferroni correction. The alpha level was set at .05 for all analyses and mean results are presented ± the standard error (SE).

Perception accuracy PA mostly highlights the possible existence of an alteration of relevant VI for anticipation. However, PA does not provide quantitative information about the extent of trajectory alteration. To quantify this trajectory alteration, we compared joint trajectories computed with the three animation methods (expressed relatively to the pelvis, to normalize global position and orientation). The trajectories were also centered to their mean value to eliminate potential offset biases. In the remaining of the paper, we focus this analysis on the hand that performs the punch. We also consider the corresponding motion capture data as the reference punching hand trajectory. Figure 3 illustrates a typical example of such a punching hand trajectory during a rear straight punch, for the three animation methods. We then compute the time-to-time distances between the reference punching hand trajectory and those obtained with the retargeting and the AMP animation methods. Based on this time-to-time distance, a histogram of errors can be computed, by grouping errors into 2 cm bins. For example, the histogram bar associated with 0 to 2 cm error, for the retargeting animation method $bar_{0-2}^{retarget}(T_i, T_f)$, is given by:

$$\frac{1}{N_{T_i,T_f}} Count_{0-2}(Abs(Hand_x^{retarget}(T_i : T_f) - Hand_x^{mocap}(T_i : T_f)) \tag{1}$$

where N_{T_i,T_f} is the number of samples in the studied time interval, $Hand_x$ is the vector of x positions of the punching hand during the interval $[T_i : T_f]$, for the three possible animation methods ($retarget$, $mocap$ or AMP), $Abs(.)$ returns the vector of absolute values of a vector, and $Count_{0-2}(.)$ returns the number of values within the interval $0 - 2$ cm.

4 Results

The perceptual accuracy for all types of punch (rear/lead straight/hook punches) and animation methods (motion capture, retargeting, AMP) are reported in Table 2 and Figure 4. The perceptual accuracy $PA_{*,*}$ ($*$ stands for all the types of punch, and animation methods) decreased as occlusion occurred earlier: $PA_{*,*}(T_s - T_1) = 0.903 \pm 0.015$, $PA_{*,*}(T_s - T_2) = 0.509 \pm 0.029$, $PA_{*,*}(T_s - T_3) =$

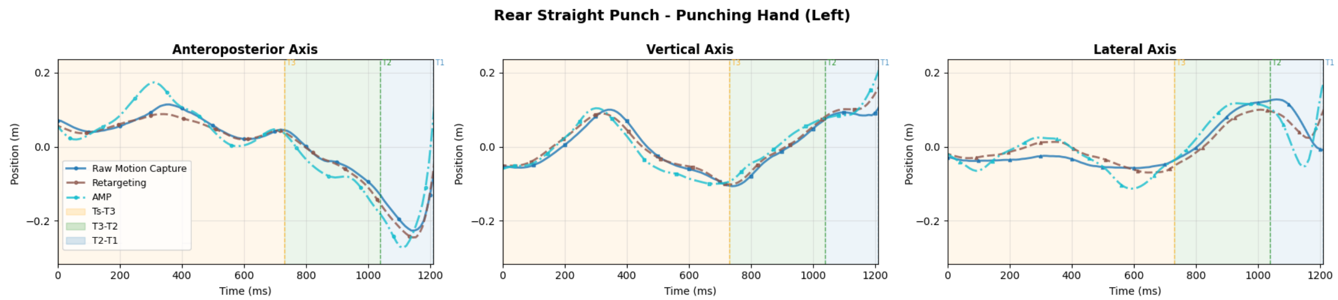

Figure 3: Trajectory of the left hand, relative to the pelvis for the three animation methods of a rear straight punch.

0.308 ± 0.026). For all the types of punch, the perceptual accuracy was $PA_{*,mocap} = 0.625 \pm 0.032$, $PA_{*,retarget} = 0.540 \pm 0.033$ and $PA_{*,AMP} = 0.555 \pm 0.033$ for the motion capture, retargeting and AMP animation methods respectively. It is important to notice that a lower PA for retargeting or AMP conditions compared to the reference motion capture condition indicates that the participants were not able to detect the relevant perceptual information. However, it does not provide any quantitative information on the extent of this potential loss of perceptual information; small or large kinematic changes could lead to the same perceptual accuracy.

Although participants had to choose among 4 different types of punch, the results show that they were more tempted to recognize straight punches compared to hooks. Among 2880 total responses, 985 were reported as lead straights (536 correct and 449 incorrect), 925 as rear straights (475 correct and 450 incorrect), 477 as lead hooks (313 correct and 164 incorrect) and 475 as rear hooks (328 correct and 147 incorrect).

4.1 Comparison to chance level

T-tests comparing each combination of condition to chance (presented with stars in Figure 4) revealed that PA was statistically above chance for all the animation methods and the type of punch $PA_{*,*}(T_s - T_1)$ (p_s < .001, except $PA_{lead\_straight,mocap}$ $(T_s - T_3)$: p = .003). For $PA_{*,*}(T_s - T_2)$ and $PA_{*,*}(T_s - T_3)$, the results show different patterns according to the type of punch. For both hooks, PA dropped below chance in these time intervals. $PA_{lead\_hook,mocap}(T_s - T_2)$ was nearly above chance level, but the p-value did not reach significance after the Bonferroni correction. $PA_{straight,*}(T_s - T_2)$ is statistically above chance (p_s from p < .005 to p < .001). $PA_{rear\_straight,*}(T_s - T_3)$ is not significantly above chance after Bonferroni correction, regardless of the animation method. For the lead straight, only the $PA_{lead\_straight,retargeting}(T_s - T_3)$ was not significantly above chance (motion capture p = .0001, AMP p = .03). As $PA_{*}(T_s - T_2)$ and $PA_{*}(T_s - T_3)$ for hooks were not significantly above chance, analyzing the statistical differences between the different animation methods would be debatable. We consequently decided to focus the remaining statistical tests on the straight punches only. Mean Cohen's d was 2.29, suggesting an overall large effect when comparing the results to chance, despite the little number of participants.

4.2 More detailed straight punches analysis

PA for straight punches was analyzed using a Generalized Linear Mixed Model (GLMM) with a binomial distribution and a logit link function. The model explained a substantial proportion of variance in accuracy (AIC = 1528.28, conditional R^2 = .276). All fixed effects coefficients and Odds ratios are presented in table 3, appendix A.5.

4.2.1 Main effects. A significant main effect of occlusion time was observed ($\chi^2(2)$ = 126.59, p < .001). Estimated PA (prob) decreased progressively as occlusion occurred earlier, corresponding to reduced vi, from prob = .876 for $[T_s - T_1]$ to prob = .768 for $[T_s - T_2]$ and prob = .510 for $[T_s - T_3]$. All pairwise comparisons were significant after Bonferroni correction (p_s < .001).

A significant main effect of the animation method was also found ($\chi^2(2)$ = 13.76, p = .001). The motion capture yielded the highest estimated PA (.802), followed by retargeting (.741) and AMP (.677). Post-hoc comparisons revealed that estimated PA for AMP differed significantly from motion capture (p < .001), whereas AMP vs. retargeting (p = .157) and motion capture vs. retargeting (p = .135) did not reach significance after Bonferroni correction.

A significant main effect of type of punch was observed ($\chi^2(1)$ = 14.50, p < .001), with higher estimated PA for Lead straight (.779) than for Rear straight (.705, p = .005).

4.2.2 Two-way interaction. The two-way interaction between occlusion time and type of punch was significant ($\chi^2(2)$ = 7.92, p = .019). Decomposing the occlusion time × type of punch interaction revealed that estimated $PA(T_s - T_1)$ did not differ between rear straight (prob = .878) and lead straight (prob = .875, p = 1.000). Estimated $PA(T_s - T_2)$ did not differ significantly between the two types of punch (rear straight: prob = .746; lead straight: prob = .788; p = 1.000). However, estimated $PA(T_s - T_3)$ for lead straight (.626) yielded significantly higher than the rear straight (.394, p < .001).

4.2.3 Three-way interaction: animation methods × occlusions × type of punch. The significant three-way interaction ($\chi^2(4)$ = 18.68, p < .001) was decomposed by comparing animation method at each combination of occlusion time and type of punch. For rear straight, no significant difference between animation method was observed. Estimated $PA(T_s - T_1)$ was comparable across AMP (.816), retargeting (.829), and motion capture (.946), with no significant pairwise differences after Bonferroni correction (AMP vs. motion capture: p = .223; AMP vs. retargeting: p = 1; motion capture vs. retargeting: p = .356). For estimated $PA(T_s - T_2)$, a similar pattern emerged (AMP:

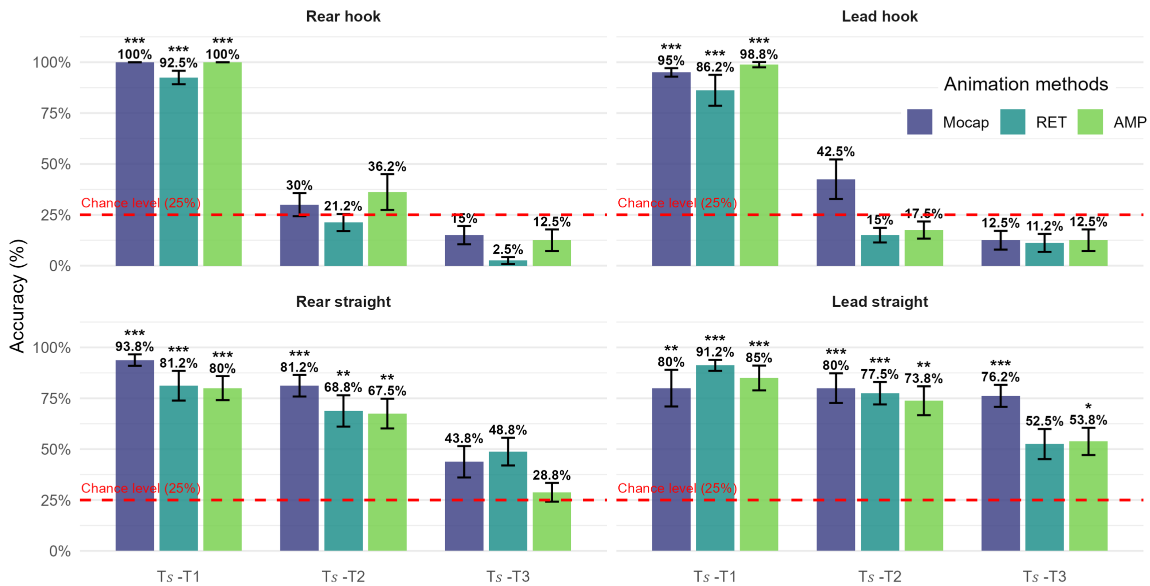

Figure 4: Mean perceptual accuracy (± SE) by occlusion times ($[T_s - T_1]$, $[T_s - T_2]$, $[T_s - T_3]$), types of punch and animation methods (Mocap: motion capure; RET: retargeting; AMP: Adversarial Motion Prior). The red dashed line represents the chance level fixed at 25%. * p < .05 ** p < .01 *** p < .001 (Bonferroni) vs chance (25%).

prob = .689; motion capture: prob = .829; retargeting: prob = .702), with no significant difference (AMP vs. motion capture: p = .752; AMP vs. retargeting: p = 1; motion capture vs. retargeting: p = 1). Estimated $PA(T_s - T_3)$ dropped markedly for AMP (.273) relative to motion capture (.433) and retargeting (.488), though pairwise comparisons did not reach significance after correction (AMP vs. motion capture: p = .738; AMP vs. retargeting: p = .132; motion capture vs. retargeting: p = 1). However, before Bonferroni correction, AMP vs. motion capture was significant (uncorrected p = .041) as AMP vs. retargeting (uncorrected p = .007).

For lead straight, PA was similar for the two animation methods, for the $[T_s - T_1]$ and $[T_s - T_2]$ intervals. Estimated $PA(T_s - T_1)$ was high across all motion types (AMP: prob = .865; motion capture: prob = .816; retargeting: prob = .923), with no significant pairwise differences (all $p_s \geq .800$). Estimated $PA(T_s - T_2)$ remained similarly high (AMP: prob = .754; motion capture: prob = .816; retargeting: prob = .791), with no significant comparisons (all $p_s = 1$). For estimated $PA(T_s - T_3)$, however, a clear differentiation emerged: the motion capture (.779) yielded significantly higher PA than AMP (.542, p = .041). The motion capture also outperformed retargeting (.529, p = .024), while AMP and retargeting did not differ (p = 1).

4.3 Kinematic analysis

Qualitative comparison of the punching hand trajectories between the three animation methods highlighted that the retargeted motion generally follows the motion capture pattern, with a mostly stable offset. To suppress this offset (partly due to difference in skeleton dimension), all the motion curves have been centered around their mean value (see Figure 3). Some high frequency information of the motion capture data seem also to be filtered by the retargeting process. We also noticed that AMP introduced some oscillations, particularly along the anteroposterior and lateral axes. Figure 5 depicts the normalized (0-100%) histogram of errors between the reference motion capture punching hand trajectory and those obtained with the retargeting and the AMP methods (as detailed in section 3.4.2). This histogram was computed using the centered trajectories, as shown in Figure 3. The histogram shows a wider spread of errors for AMP compared to retargeting, for all the axes, with errors going up to 1.8 cm. 0 to 2 cm errors also occur less times when using AMP (28 to 78% for the anteroposterior and lateral axes) compared to retargeting (60 to 100% for the anteroposterior and lateral axes). The punching hand movement seems to be less accurately animated for the lateral axis compared to the two other axes, for both retargeting and AMP conditions. The animation errors for the $[T_3 - T_2]$ interval seem to be less scattered, with very little to no error above 2 cm (resp. 8 cm) for the retargeting (resp. AMP) method. For the first time interval $[T_s - T_3]$ and the last one $[T_2 - T_1]$, important errors (> 8 cm) more frequently occurred using AMP compared to retargeting.

5 Discussion

5.1 Perceptual accuracy over different occlusion times

As expected, the perceptual accuracy naturally declined with the decrease of the quantity of available information: shorter motion sequences leaded to less PA compared to longer sequences. For the longest sequences, $PA_{*,*}(T_s - T_1)$ was 0.903 for all punches; participants were able to identify the punch almost 90% of the time. In these long sequences, a large amount of VI is available, and the punching arm has already started to move to the target. $PA_{straight,*}(T_s - T_2)$ slightly decreased down to 67-81% for the

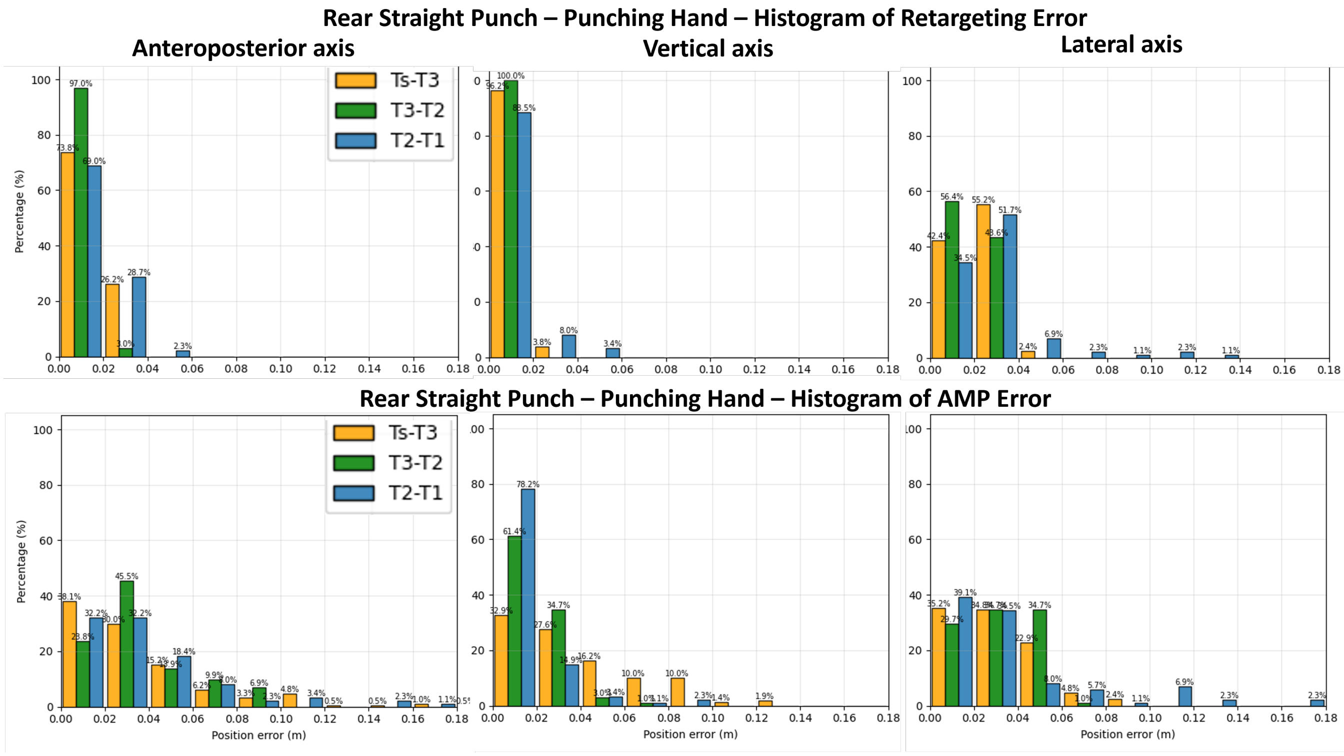

Figure 5: Histogram of errors between the punching hand motion capture trajectory and those computed by the retargeting and the AMP animation methods.

straight punches, which remains above chance. This small decrease of performance support the hypothesis that the expert boxers found enough VI before the punching arm moves to the target, for this type of punch.

In contrast, $PA_{hooks,*}(T_s - T_2)$ collapsed below chance level, for hooks, which were statistically less frequently selected compared to straight punches. For $[T_s - T_3]$, which ends at the initiation of the last support leg's extension, only lead straights were detected above chance. Let us recall the unbalanced distribution of participants' responses, with a clear preference for straight punches. It resulted in a significantly higher number of wrong straight punch responses compared to the wrong hook responses. When participants responded with a hook, it tended to be more accurate. Thus, there is a potential overestimation of the straight punch's PA. In the context of French boxing, straight punches are more common than hooks, which are often incorporated into a sequence. Consequently, participants may have responded more often with "straight" than "hook" due to the prevalence of this punch in isolated condition. It would be interesting to test if a group of Olympic boxing participants would lead to a different result.

Appendix A.3 provide information about the speed for each type of punch. Surprisingly, the participants obtained the highest perceptual score for lead straight punch, with the shorter duration and higher speed. This result suggests complex influence of speed and other kinematic parameters, that should be more deeply analyzed.

5.2 Impact of the animation methods

Overall, PA was better in the motion capture condition than in the retargeting and AMP conditions. However, this effect varied depending on the combination of conditions. Since PA for hooks rapidly fell below chance level, hook trials were excluded from the statistical analysis. Focusing on straight punches, the main effect indicated that PA for AMP was significantly lower than for motion capture. It also indicates that retargeting was close to AMP. This effect was particularly true for the lead straight in $[T_s - T_3]$, where the motion capture was above the two other conditions.

Retargeting, in this simple implementation, does not solve complex constraints, compared to more recent works. Despite this little adaptation of the original motion capture data, some original oscillations disappear (see Figure 3, confirmed by a Spectral analysis in Appendix A.4). This kind of modification may disrupt the VI used for anticipation. Although AMP was trained on these retargeted motions, with discarded oscillations, it introduces new oscillations with different frequency (see Figure 9), visible in Figure 3 for the $[T_s - T_1]$ interval and the anteroposterior axis. In this case, the added oscillation went up to 20cm, creating misleading information confusing the prediction of the punch. Such motion artifacts may be interpreted by expert boxers as a feint or as a preparatory signal for an upcoming punch. This is consistent with the drop of perception accuracy observed at $[T_s - T_3]$ for the rear straight punch, consistent with substantial animation errors (up to 12 cm) reported during the same interval (see histogram in Figure 5). For the lead straight punch, these errors may also account for the decrease of PA in $[T_s - T_3]$, for retargeting and AMP conditions compared to motion capture. Overall, these results support the hypothesis that animation methods alter the VI required to anticipate. Since PA for the rear straight punch falls below chance level, a similar analysis is not meaningful.

In this paper, we used the retargeting method commonly used in physics-based imitation learning [Peng et al. 2021]. However, motion retargeting is a very active research area, with many recent innovative deep learning–based approaches. It would be interesting to benchmark these recent alternative methods. However, most of these works focus on contact preservation and penetration avoidance, which are not critical for the boxing motions studied here. In addition, most of the previous works rely on replaying the original joint angles with adaptations to mitigate artifacts. Even in the case of contact-free and collision-free motions, such methods can still introduce subtle changes to the original motion, with no guarantee of preserving the relevant VI. This notion of "relevant" VI is difficult to define, model, quantify, and might be application-dependent. Further work is therefore needed to better understand which intrinsic properties of motion should be preserved when transferring motion capture data to a virtual human.

5.3 Limitations

Ten French boxers participated in the experiment, and extending this sample would improve the statistical power of the analysis. As reported above, the results may change depending on the speciality of the boxing participants. It would therefore be valuable to include practitioners from a broader range of boxing disciplines to validate this assumption, with potential implications for the design of dedicated VR training simulators. More generally, this approach could be extended to other interactive contexts in which individuals anticipate the actions of others.

We initially planned to place a camera to use videos as reference. However, in the specific context of boxing, this video is very difficult to get without occlusions (hand protecting against the punches). Moreover, it would also raise the question of comparing video and computer generated sequences.

Motion capture data was collected from a single boxer with a specific individual style. The duration of the perceptual study was very long, combining three animation methods, four different punches, with several repetitions. It is important that the participants react to the same style for all these combination, to avoid taking into account the effect of boxing style. Adding new boxers would consequently significantly increase the number of trials to evaluate. Future work should include multiple boxers to assess whether animation methods consistently affect different motion styles.

Recent work motion encoding concept [Becchio et al. 2024] could help us to identify the most relevant kinematic features that encode the type of punch. However, we had a little dataset of examples, which is too limited to carry-out extensive and robust statistical analysis about the link between kinematic variables and the type of punch. Consequently, the aim of this paper was to demonstrate that computer animation frameworks can disrupt important kinematic information (without being able to identify which one), altering the decision-making of experts used to anticipate this type of motion.

Moreover, previous research has shown that perceptual studies do not fully capture information processing in complex sport environments [Mann et al. 2010]. In particular, the absence of a perception–action coupling is a limitation of the present study. It would therefore be interesting to evaluate whether punch prediction accuracy remains consistent in interactive tasks involving action, for instance using VR systems in which participants both perceive and respond to simulated movements.

In this paper, AMP was trained with a SMPL skeleton, requiring retargeting before training. Hence, we can report how errors accumulate as AMP is used, creating unrealistic oscillations certainly due to the PD controllers. To compare separately retargeting and AMP, we could train AMP with the motion capture skeleton, but it should be validated properly first, as in previous AMP-like works.

This work assumes that motion capture data provide a reliable reference condition, although they resulted from tracking a skeleton motion from marker positions, which also may introduce errors (such as soft tissue artefacts [Ozan et al. 2026]). One alternative might be to film the opponent from a first-person perspective. In practice, this is difficult to achieve in boxing scenarios, where an opponent often holds pads that obstruct the camera viewpoint.

6 Conclusion

In this paper, we investigated how state-of-the-art virtual human animation methods can disrupt kinematic information commonly used by experts to anticipate an opponent's action in adversarial scenarios. In a boxing context, we employed an occlusion paradigm to evaluate expert boxers' ability to anticipate punch types from incomplete animation sequences generated using motion capture, motion retargeting, and physics-based reinforcement learning (AMP) methods. The results show better perceptual accuracy for motion capture sequences compared to those produced by motion retargeting and AMP. This difference is more pronounced for short sequences, where only early visual information—critical for anticipation—is available. An analysis of the effector trajectories across animation methods further confirms that retargeting and AMP introduce alterations relative to motion capture. Overall, these findings indicate that commonly used animation methods, while designed to reproduce human motion, can introduce distortions that may be a significant limitation when designing VR simulators for expert users. This issue is crucial when aiming to maximize the transfer of skills learned in VR to real-world practice. It also opens new research directions toward designing animation methods based on new metrics that better preserve the VI required for natural expert perception and interaction with virtual humans.

Acknowledgments

This work is part of REVEA, DIGISPORT, PEPR eNSEMBLE and CONTINUUM projects supported by the ANR in the "France 2030" program (20-STHP-0004, ANR-18-EURE-0022, ANR-22-EXEN-0004 and ANR-21-ESRE-0030). Experiments were carried out using the Grid'5000 testbed, supported by Inria, CNRS, RENATER and several Universities (https://www.grid5000.fr).

References

Jinseok Bae, Younghwan Lee, Donggeun Lim, and Young Min Kim. 2025. PLT: Part-Wise Latent Tokens as Adaptable Motion Priors for Physically Simulated Characters.

In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*. 1–10.
Jean Basset, Stefanie Wuhrer, Edmond Boyer, and Franck Multon. 2020. Contact preserving shape transfer: Retargeting motion from one shape to another. *Computers & Graphics* 89 (2020), 11–23.
Cristina Becchio, Kiri Pullar, Eugenio Scaliti, and Stefano Panzeri. 2024. Kinematic coding: Measuring information in naturalistic behaviour. *Physics of Life Reviews* 51 (2024), 442–458. doi:10.1016/j.plrev.2024.11.009
Benoit Bideau, Richard Kulpa, Nicolas Vignais, Sébastien Brault, Franck Multon, and Cathy Craig. 2010. Using virtual reality to analyze sports performance. *IEEE Comput Graph Appl* 30, 2 (April 2010), 14–21.
T. Cheynel, T. Rossi, B. Bellot-Gurlet, D. Rohmer, and M.P. Cani. 2025. ReConForM : Real-time Contact-aware Motion Retargeting for more Diverse Character Morphologies. *Computer Graphics Forum* 44, 2 (2025), e70028. doi:10.1111/cgf.70028
J. Gibson. 1979. *The Ecological Approach to Perception.* Boston : Haughton Mifflin.
Michael Gleicher. 1998. Retargetting motion to new characters. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques.* 33–42.
Thomas Hausegger, Christian Vater, and Ernst-Joachim Hossner. 2019. Peripheral Vision in Martial Arts Experts: The Cost-Dependent Anchoring of Gaze. *Journal of Sport and Exercise Psychology* 41, 3 (June 2019), 137–146.
Edmond SL Ho, Taku Komura, and Chiew-Lan Tai. 2010. Spatial relationship preserving character motion adaptation. In *ACM SIGGRAPH 2010 papers*. 1–8.
Jonathan Ho and Stefano Ermon. 2016. Generative adversarial imitation learning. *Advances in neural information processing systems* 29 (2016).
Inseo Jang, Soojin Choi, Seokhyeon Hong, Chaelin Kim, and Junyong Noh. 2024. Geometry-aware retargeting for two-skinned characters interaction. *ACM Transactions on Graphics (TOG)* 43, 6 (2024), 1–17.
Taeil Jin, Meekyoung Kim, and Sung-Hee Lee. 2018. Aura mesh: Motion retargeting to preserve the spatial relationships between skinned characters. In *Computer Graphics Forum*, Vol. 37. Wiley Online Library, 311–320.
Sophie Kenny, Naureen Mahmood, Claire Honda, Michael J. Black, and Nikolaus F. Troje. 2019. Perceptual Effects of Inconsistency in Human Animations. *ACM Trans. Appl. Percept.* 16, 1 (2019), 18 pages. doi:10.1145/3301411
Richard Kulpa, Franck Multon, and Bruno Arnaldi. 2005. Morphology-independent representation of motions for interactive human-like animation. In *Eurographics*, European Association for Computer Graphics (Ed.). M. Alexa, J. Marks, Dublin,/Ireland, France. https://inria.hal.science/inria-00000220 http://www.eg.org/.
Yoonsang Lee, Sungeun Kim, and Jehee Lee. 2010. Data-driven biped control. In *ACM SIGGRAPH 2010 papers*. 1–8.
Libin Liu, Michiel Van De Panne, and KangKang Yin. 2016. Guided learning of control graphs for physics-based characters. *ACM Transactions on Graphics (TOG)* 35, 3 (2016), 1–14.
Zhengyi Luo, Jinkun Cao, Josh Merel, Alexander Winkler, Jing Huang, Kris Kitani, and Weipeng Xu. 2024. Universal Humanoid Motion Representations for Physics-Based Control. doi:10.48550/arXiv.2310.04582 arXiv:2310.04582 [cs].
Zhengyi Luo, Jinkun Cao, Alexander Winkler, Kris Kitani, and Weipeng Xu. 2023. Perpetual Humanoid Control for Real-time Simulated Avatars. doi:10.48550/arXiv.2305.06456 arXiv:2305.06456 [cs].
David Mann, A. Mark Williams, Paul Ward, and Christopher M. Janelle. 2007. Perceptual-cognitive expertise in sport: a meta-analysis. *Journal of Sport & Exercise Psychology* 29, 4 (Aug. 2007), 457–478.
David L. Mann, Bruce Abernethy, and Damian Farrow. 2010. Action specificity increases anticipatory performance and the expert advantage in natural interceptive tasks. *Acta Psychologica* 135, 1 (2010), 17–23. doi:10.1016/j.actpsy.2010.04.006
Stefan C. Michalski, Ancret Szpak, and Tobias Loetscher. 2019. Using Virtual Environments to Improve Real-World Motor Skills in Sports: A Systematic Review. *Frontiers in Psychology* 10 (Sept. 2019), 2159. doi:10.3389/fpsyg.2019.02159
Sean Müller, Evan Dekker, Khaya Morris-Binelli, Benjamin Piggott, Gerard Hoyne, Wayne Christensen, Peter Fadde, Leonard Zaichkowsky, John Brenton, and David Z. Hambrick. 2022. Attributes of Expert Anticipation Should Inform the Design of Virtual Reality Simulators to Accelerate Learning and Transfer of Skill. *Sports Medicine* (July 2022). doi:10.1007/s40279-022-01735-7
Simon Ozan, Loïc Fourel, Pierre Touzard, Richard Kulpa, and Caroline Martin. 2026. Uncertainty estimation in marker-based motion capture of the tennis serve. *Journal of Biomechanics* 195 (2026), 113119. doi:10.1016/j.jbiomech.2025.113119
Liang Pan, Zeshi Yang, Zhiyang Dou, Wenjia Wang, Buzhen Huang, Bo Dai, Taku Komura, and Jingbo Wang. 2025. TokenHSI: Unified Synthesis of Physical Human-Scene Interactions through Task Tokenization. doi:10.48550/arXiv.2503.19901 arXiv:2503.19901 [cs].
Jonathan Peirce, Jeremy Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Lindeløv. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* 51 (02 2019). doi:10.3758/s13428-018-01193-y
Xue Bin Peng, Pieter Abbeel, Sergey Levine, and Michiel Van de Panne. 2018. Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions On Graphics (TOG)* 37, 4 (2018), 1–14.
Xue Bin Peng, Yunrong Guo, Lina Halper, Sergey Levine, and Sanja Fidler. 2022. ASE: Large-Scale Reusable Adversarial Skill Embeddings for Physically Simulated Characters. (2022). doi:10.48550/ARXIV.2205.01906
Xue Bin Peng, Ze Ma, Pieter Abbeel, Sergey Levine, and Angjoo Kanazawa. 2021. Amp: Adversarial motion priors for stylized physics-based character control. *ACM Transactions on Graphics (ToG)* 40, 4 (2021), 1–20.
Ross A Pinder, Keith Davids, Ian Renshaw, and Duarte Araújo. 2011. Representative learning design and functionality of research and practice in sport. *Journal of Sport and Exercise Psychology* 33, 1 (2011), 146–155.
Paul S. A. Reitsma and Nancy S. Pollard. 2003. Perceptual metrics for character animation: sensitivity to errors in ballistic motion. *ACM Trans. Graph.* 22, 3 (2003), 537–542. doi:10.1145/882262.882304
Thomas Romeas, Basil More-Chevalier, Mathieu Charbonneau, and François Bieuzen. 2022. Virtual-Reality Training of Elite Boxers Preparing for the Tokyo 2020 Olympics During the COVID-19 Pandemic: A Case Study. *Case Studies in Sport and Exercise Psychology* 6, 1 (Jan. 2022), 21–35. doi:10.1123/cssep.2021-0027
Tao Song, Mingzhu Ye, Gesi Teng, Weikun Zhang, and Antao Chen. 2025. The role of action anticipation in specific sport performance: A three-level meta-analysis and systematic review in temporal occlusion paradigm. *Psychology of Sport and Exercise* 79 (July 2025), 102839. doi:10.1016/j.psychsport.2025.102839
Chen Tessler, Yoni Kasten, Yunrong Guo, Shie Mannor, Gal Chechik, and Xue Bin Peng. 2023. CALM: Conditional Adversarial Latent Models for Directable Virtual Characters. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Proceedings*. ACM, Los Angeles CA USA, 1–9. doi:10.1145/3588432.3591541
A. Williams and R. Jackson. 2019. Anticipation in sport: Fifty years on, what have we learned and what research still needs to be undertaken? *Psychology of Sport and Exercise* 42 (May 2019), 16–24. doi:10.1016/j.psychsport.2018.11.014
Kaixiang Xie, Pei Xu, Sheldon Andrews, Victor B Zordan, and Paul G Kry. 2023. Too stiff, too strong, too smart: Evaluating fundamental problems with motion control policies. *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 6, 3 (2023), 1–17.
Michael Xu, Yi Shi, KangKang Yin, and Xue Bin Peng. 2025. Parc: Physics-based augmentation with reinforcement learning for character controllers. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*. 1–11.
Pei Xu and Ioannis Karamouzas. 2021. A gan-like approach for physics-based imitation learning and interactive character control. *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 4, 3 (2021), 1–22.
Goksu Yamac, Carol O'Sullivan, and Michael Neff. 2024. Understanding the Impact of Visual and Kinematic Information on the Perception of Physicality Errors. *ACM Trans. Appl. Percept.* 22, 1 (2024), 30 pages. doi:10.1145/3660636
Zijie Ye, Jia-Wei Liu, Jia Jia, Shikun Sun, and Mike Zheng Shou. 2024. Skinned motion retargeting with dense geometric interaction perception. *Advances in Neural Information Processing Systems* 37 (2024), 125907–125934.
Mohamed Younes, Ewa Kijak, Richard Kulpa, Simon Malinowski, and Franck Multon. 2023. MAAIP: Multi-Agent Adversarial Interaction Priors for imitation from fighting demonstrations for physics-based characters. (2023). doi:10.48550/ARXIV.2311.02502 Publisher: arXiv Version Number: 1.

A Appendices

In the five following appendices, we report additional information about the protocol, such as the definition of the various skeleton used for this study (see appendix A.1). We also introduce additional results to deeper analyze the perceptual performance of the participants (see appendix A.2). We added two additional information about the kinematic information embedded in the 3 animations: motion speed (appendix A.3) and spectrum (appendix A.4) analyses. Complementary details about the Generalized Linear Mixed model are given in appendix A.5, to explain how it is used to carry-out the statistical analysis of the perceptual study .

A.1 Retargeting

Retargeting is performed by establishing a consistent joint-level correspondence between the motion capture and the SMPL skeletons, ensuring that each joint from the motion capture skeleton maps to its semantically equivalent counterpart (see Figure 6). Table 1 reports the correspondence used in our retargeting stage, detailing the mapping between the 24 joints of both skeletons and their associated anatomical points.

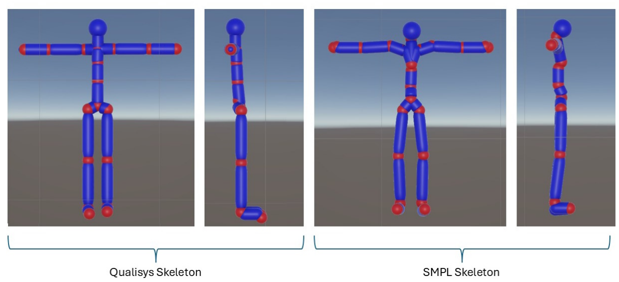

Figure 6: T-poses of the two skeletons (Qualisys and SMPL) used in the experiment. The hierarchy is shown in blue and the joints in red (for visualization purposes only; in the experiment, joints are also represented in blue).

Table 1: Joints mapping table between motion capture and SMPL skeletons with anatomical description.

# Joint	Motion capture	SMPL	Anatomical points
1	Hips	Pelvis	pelvic center
2	Spine	Spine	Lower spine, lumbar region
3	Spine1	Torso	Upper abdominal region
4	Spine2	Chest	Sternum
5	Neck	Neck	Cervical junction
6	Head	Head	Head center
7	LeftShoulder	L_Thorax	Left Sternoclavicular joint center
8	LeftArm	L_Shoulder	Left shoulder center
9	LeftForeArm	L_Elbow	Left elbow center
10	LeftForeArmRoll	L_Wrist	Left Wrist center
11	LeftHand	L_Hand	Left fingers base center
12	RightShoulder	R_Thorax	Right Sternoclavicular joint center
13	RightArm	R_Shoulder	Right shoulder center
14	RightForeArm	R_Elbow	Right elbow center
15	RightForeArmRoll	R_Wrist	Right Wrist center
16	RightHand	R_Hand	Right fingers base center
17	LeftUpLeg	L_Hip	Left hip center
18	LeftLeg	L_Knee	Left knee center
19	LeftFoot	L_Ankle	Left ankle center
20	LeftToeBase	L_Toe	Left toe base
21	RightUpLeg	R_Hip	Right hip center
22	RightLeg	R_Knee	Right knee center
23	RightFoot	R_Ankle	Right ankle center
24	RightToeBase	R_Toe	Right toe base center

A.2 Mean perceptual accuracy details

Table 2 presents the mean perceptual accuracy values for each combination of occlusion time, animation method, and type of punch. The values are presented with standard error. These values were used to create Figure 4.

Table 2: Perceptual accuracy (mean ± Standard Error) across occlusion times, types of punch, and animation methods.

Occlusion	Type of Punch	Processing stage	Accuracy (%) ± SE
$[T_s - T_1]$	Rear hook	AMP	100.0 ± 0.0
		Mocap	100.0 ± 0.0
		RET	92.5 ± 3.3
	Lead hook	AMP	98.8 ± 1.3
		Mocap	95.0 ± 2.1
		RET	86.2 ± 7.6
	Rear straight	AMP	80.0 ± 5.9
		Mocap	93.8 ± 2.8
		RET	81.2 ± 7.3
	Lead straight	AMP	85.0 ± 6.1
		Mocap	80.0 ± 9.0
		RET	91.2 ± 2.7
$[T_s - T_2]$	Rear hook	AMP	36.2 ± 8.8
		Mocap	30.0 ± 5.7
		RET	21.2 ± 4.2
	Lead hook	AMP	17.5 ± 4.2
		Mocap	42.5 ± 9.7
		RET	15.0 ± 3.6
	Rear straight	AMP	67.5 ± 7.3
		Mocap	81.2 ± 5.3
		RET	68.8 ± 7.7
	Lead straight	AMP	73.8 ± 7.1
		Mocap	80.0 ± 7.3
		RET	77.5 ± 5.5
$[T_s - T_3]$	Rear hook	AMP	12.5 ± 5.3
		Mocap	15.0 ± 4.5
		RET	2.5 ± 1.7
	Lead hook	AMP	12.5 ± 5.3
		Mocap	12.5 ± 4.6
		RET	11.2 ± 4.4
	Rear straight	AMP	28.8 ± 4.6
		Mocap	43.8 ± 7.7
		RET	48.8 ± 6.8
	Lead straight	AMP	53.8 ± 6.7
		Mocap	76.2 ± 5.4
		RET	52.5 ± 7.4

A.3 Analysis of the motion speed

In such highly dynamic motion, and consequently difficult anticipatory task, motion speed could impact the participants' answer. Especially, fast motion may be more difficult to anticipate than slow ones, and little modification in high speed motion may impact more the anticipatory skills. To evaluate the validity of this assumption, we consequently studied the velocity peak for each type of punch, and inside each segment of the motion (especially during T3-T2, T2-T1), as shown in Figure 7. Let us recall that these plots were obtained with 4 trials per condition only, making it difficult to carry-out accurate statistical analysis. During T3-T2, there was no big difference in the motion capture speed peak.

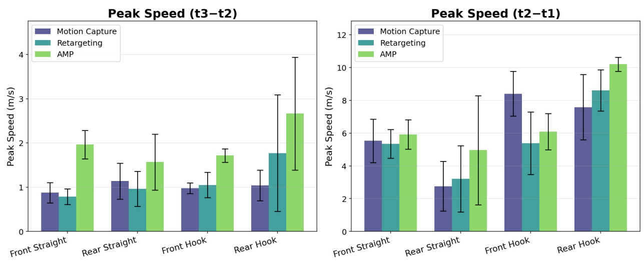

Figure 7: Peak speed of the motions in T3-T2 and T2-T1.

However, the duration of the front straight (see figure 8) punch was half the one of the other punches (the striking arm starts moving in front of the boxer, decreasing the distance and time to target). During T2-T1, just before the contact, the speed peak seems higher for the hooks compared to the straight punches. However, the duration of this segment was the same for all the punches, except for the shorter front straight punch, similarly to T3-T2. In some cases, one can notice that important peak speeds were modified using retargeting or AMP motions. If we consider the perceptual study (see figure 4), the participants obtained the highest score for lead straight punch, the one with the shorter duration. It would tend to lead to the inverse conclusion: faster motions lead to less visual perturbation. Our analysis is that this expert population used to interact with this specific lead straight punch, which is commonly used by many boxers; the most used punch in this discipline. It would be interesting to explore more this direction by artificially tuning the velocity, and by recruiting different populations with different backgrounds for the perceptual study. With the current information, it seems hazardous to formally respond to this question.

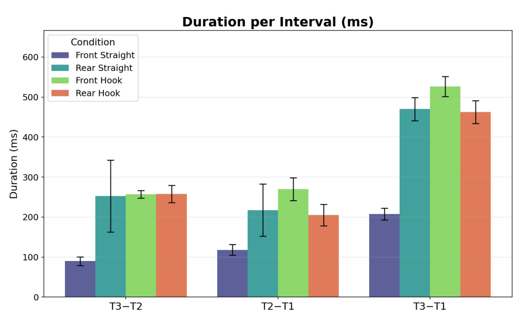

Figure 8: Duration of the intervals T3-T2 and T2-T1 for each type of motion.

A.4 Frequency domain analysis

In order to more deeply understand where the kinematic signal is being changed when using animation methods, and especially the appearance of oscillations discarded by the retargeting method, and new ones added by AMP, we performed some spectral analysis (Welch analysis, nperseg = 256, 50% overlap, Hann window). The spectrum of the retargeted motions exhibit a slight overall attenuation, with occasional local losses of energy (see Figure 9, 2–5 Hz on the lateral axis). In contrast, AMP tends to introduce additional energy in the 5–12 Hz range, which is consistent with oscillations likely originating from the proportional–derivative controller used in the simulation. Reversely, the 1.5-2.5Hz natural oscillations were filtered by the retargeting method.

A.5 Generalized Linear Mixed model's details

Table 3 reports the Fixed Effects Coefficients and Odds Ratios from the Binomial GLMM, used to evaluate the significance of the statistical results with our sample size.

The fixed effects coefficients (β) represent the change in log-odds of a correct response associated with each predictor relative to the reference categories (Occlusion Time = T1; Animation Method = AMP; Punch Type = Rear straight), while odds ratios (OR) express these effects on a multiplicative scale, indicating how much more - or less - likely a correct response is in one condition compared to the reference. Reporting both metrics is relevant as they provide complementary information: β values allow assessment of the direction and statistical significance of each effect within the logistic framework, whereas ORs offer an intuitive measure of effect size that facilitates direct comparison across conditions and interpretation of interaction terms. Overall, occlusion time was the dominant factor shaping perceptual accuracy, with a large and systematic decrease in the odds of correct responding from T1 to T3 (OR = 0.09), indicating that maximal occlusion reduced the likelihood of a correct response by over 90% relative to minimal occlusion. Strikingly, the three-way interaction revealed that motion capture animations under maximal occlusion and lead straight punches multiplied the odds of correct responding by more than eightfold compared to AMP motion (OR = 8.22), suggesting a practically large advantage of high-fidelity motion capture in preserving perceptual discriminability when VI is severely degraded.

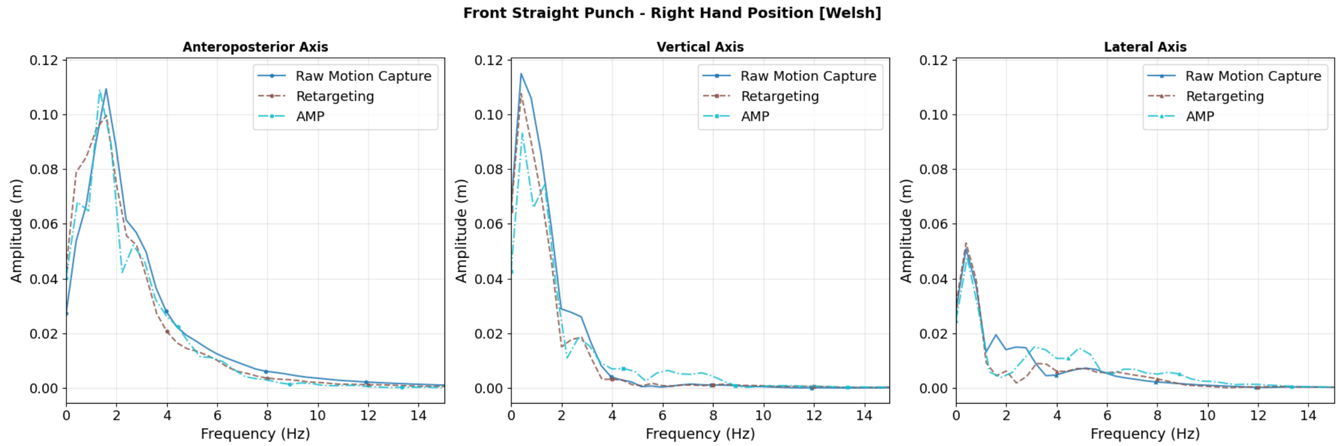

Figure 9: Frequency domain analysis of the position of the right hand of a front straight punch (hitting hand).

Table 3: Fixed Effects Coefficients and Odds Ratios from the Binomial GLMM

Predictor	β	SE	OR
Main effects			
(Intercept)	**1.49**	**0.344**	**4.44**
OcclusionTime T2	−0.70	0.378	0.50
OcclusionTime T3	**−2.47**	**0.386**	**0.09**
Animation Mocap	**1.36**	**0.546**	**3.91**
Animation RET	0.08	0.409	1.09
PunchType lead straight	0.36	0.428	1.44
Two-way interactions			
OcclusionTime T2 × Animation Mocap	−0.59	0.666	0.56
OcclusionTime T3 × Animation Mocap	−0.65	0.647	0.52
OcclusionTime T2 × Animation RET	−0.02	0.539	0.98
OcclusionTime T3 × Animation RET	0.85	0.536	2.33
OcclusionTime T2 × PunchType lead straight	−0.04	0.559	0.96
OcclusionTime T3 × PunchType lead straight	0.78	0.551	2.19
Animation Mocap × PunchType lead straight	**−1.73**	**0.693**	**0.18**
Animation RET × PunchType lead straight	0.55	0.654	1.72
Three-way interactions			
OcclusionTime T2 × Animation Mocap × PunchType lead straight	1.32	0.881	3.75
OcclusionTime T3 × Animation Mocap × PunchType lead straight	**2.11**	**0.854**	**8.22**
OcclusionTime T2 × Animation RET × PunchType lead straight	−0.39	0.833	0.68
OcclusionTime T3 × Animation RET × PunchType lead straight	−1.53	0.811	0.22

Note. β = unstandardized log-odds coefficient; SE = standard error; OR = odds ratio. Reference categories: Occlusion Time = T1; Animation method = AMP; Punch Type = Rear straight. Significant predictors are shown in **bold**.