Too Stiff, Too Strong, Too Smart: Evaluating Fundamental Problems with Motion Control Policies

KAIXIANG XIE, McGill University, Canada

PEI XU, Clemson University, USA

SHELDON ANDREWS, École de technologie supérieure, Canada / Roblox, USA

VICTOR B. ZORDAN, Clemson University, USA / Roblox, USA

PAUL G. KRY, McGill University, Canada

Deep reinforcement learning (DRL) methods have demonstrated impressive results for skilled motion synthesis of physically based characters, and while these methods perform well in terms of tracking reference motions or achieving complex tasks, several concerns arise when evaluating the naturalness of the motion. In this paper, we conduct a preliminary study of specific quantitative metrics for measuring the naturalness of motion produced by DRL control policies beyond their visual appearance. Namely, we propose to study the stiffness of the control policy, in anticipation that it will influence how the character behaves in the presence of external perturbation. Second, we establish two baselines for strength that allow evaluating the use of joint torques in comparison to human performance. Third, we propose the study of variability to reveal the unnatural precision of control policies and how they compare to real human motion. In sum, we aim to establish repeatable measures to assess the naturalness of control policies produced by DRL methods, and we present a set of comparisons from state-of-the-art systems. Finally, we propose simple modifications to improve realism on these axes.

Additional Key Words and Phrases: computer animation, motion control, reinforcement learning

1 INTRODUCTION

Over the last decade, deep reinforcement learning (DRL) methods have become widely adopted and are now recognized as versatile for a wide range of behaviors in physically based character animation. With few exceptions, these approaches rely heavily on the use of motion capture clips as reference data in order to generate natural-looking motion. Often through direct imitation or by learning generative models from sets of data, today's controllers can imitate human motion quite well, meaning they appear *kinematically* similar to motion capture references. However, a key benefit of these physical agents is their ability to deviate from kinematic playback of data to extend their applicability to different environments and unexpected perturbations. In this capacity, many of the state-of-the-art techniques quickly reveal shortcomings in the learned control policies,

Authors' addresses: Kaixiang Xie, McGill University, Canada, kaixiang.xie@mail.mcgill.ca; Pei Xu, Clemson University, USA, peix@clemson.edu; Sheldon Andrews, École de technologie supérieure, Canada / and Roblox, USA, sheldon.andrews@etsmtl.ca; Victor B. Zordan, Clemson University, USA / and Roblox, USA, vbzordan@roblox.com; Paul G. Kry, McGill University, Canada, kry@cs.mcgill.ca.
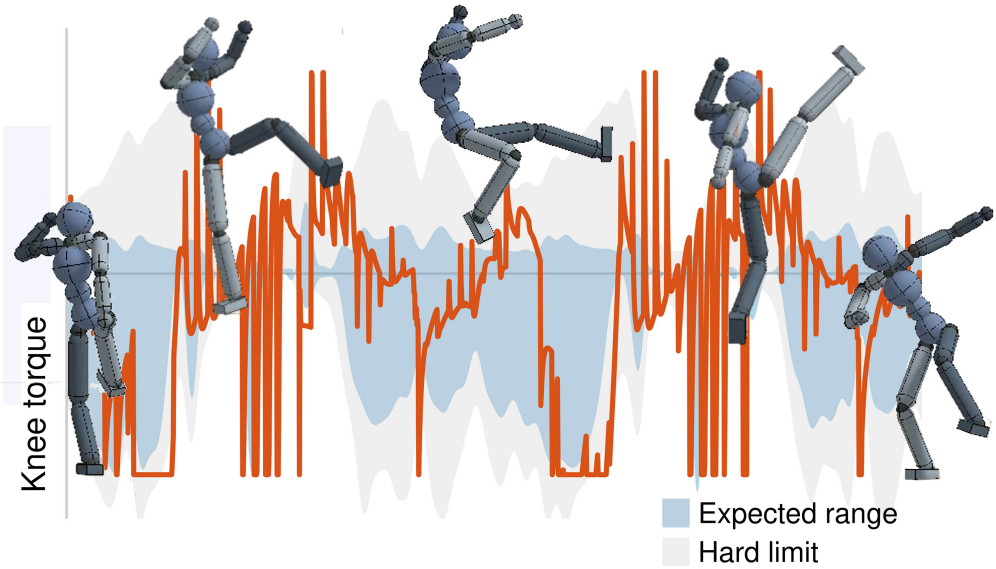
Fig. 1. The orange curve shows knee torques generated by a DeepMimic policy. The blue area shows the expected range of knee torque using our data-driven approach. The grey area shows a torque limit computed using the biomechanics model [Anderson et al. 2007]. The DeepMimic controller frequently outputs torques that are outside of the expected range. Sometimes, these torques also violate the hard limit.

for example, with highly unnatural interactions between the characters and colliding objects, or impacts with the environment.

We posit that in order to move past the current shortcomings, the community needs established metrics and baselines that go beyond kinematic comparisons (e.g., joint angles or end-effector trajectory comparisons). To this end, this paper presents a set of quantitative comparisons between the output produced by recent DRL methods and real human data and biomechanical models. To move past the visual appearance, we evaluate naturalness by proposed measures that are not directly kinematic, and attempt to tease out how similar (or different) their output is from humans acting in similar ways. Further, we strive to propose metrics that will indicate areas for improvement and extended investigation. Due to their availability and relevance, we primarily compare policies trained using DeepMimic (DM) [Peng et al. 2018a], Adversarial Motion Priors (AMP) [Peng et al. 2021], and Interactive Character Control Generative Adversarial Network (ICCGAN) [Xu and Karamouzas 2021], and evaluate their naturalness along several axes. In support of our hypothesis, we find that while these control policies do a great job at imitating human motions (albeit their intended goal), there is still a significant difference between the controlled characters and their real-world counterparts. Our preliminary study shows that the control policies have shortcomings in the naturalness of the motion they generate. More specifically, we introduce three new metrics that evaluate the naturalness of control policies from the following: stiffness, measured by the computed stiffness of the control policy; strength, comparing control torques in comparison to the range of joint torques found in humans, and "smarts", assessed by variability in cyclic motions in contrast to that seen in humans. The latter aims to reveals how controllers are in general too precise and repeatable (robotic), and therefore not humanlike. Beyond these proposed metrics,

we also suggest modifications that aim to improve motion naturalness based on our evaluation and showcase the impacts of these changes through external perturbations applied to controlled characters.

Notably, as our key contribution is to propose grounded metrics that reveal differences that are beyond the kinematic "mask" of natural appearance, the characteristics of human motion presented in this paper are largely indiscernible simply from viewing the resulting animations. Instead, we focus on careful analysis of motion data using the observations and plots presented in this paper. Therefore, even though this is an animation paper, we forego the traditional animation playback of results, and lean on the visualizations presented in the text. The supplemental documents animations for reference, however the components of the embedded analysis are, by their nature, not easy to see in the resulting animation.

2 RELATED WORK

This section briefly summarizes the body of work on animating physics-based 3D characters using RL. For a more complete overview of the topic, we direct the reader to the recent surveys by Mourot et al. [2022] and Kwiatkowski et al. [2022]. Since much of our analyses focuses on evaluating character animations with respect to the biological limitations of human motion, related work in biomechanics and neuroscience is also discussed, along with related methods in computer graphics.

2.1 Physics-based Motion Synthesis

Typical approaches for learning control policies for human motion synthesis construct a reward function based on imitation objectives that attempt to minimize the differences between the motion produced by the simulated character and an ensemble of reference clips [Bergamin et al. 2019; Peng et al. 2018a] or videos [Peng et al. 2018b]. However, such approaches require careful selection of the clips to be imitated, and a high-level motion planner or policy must be trained to determine transitions between clips for performing general tasks. Other work avoids the issue of controller sequencing by learning a generative motion-prior to synthesize motion that is similar to those found in large, unstructured motion datasets [Peng et al. 2021; Xu and Karamouzas 2021]. These methods do not explicitly track individual motion clips, and thus are capable of producing a wider variety of motions. However, our experiments demonstrate that they do not exhibit certain characteristics of real human motion. While the focus of our work is on evaluating the naturalness of motion produced by control policies, other recent studies focus on evaluating their stability and robustness [Park et al. 2022].

Compliance is a mechanism exploited by humans for successful control [Bennett et al. 1992]. In computer graphics, previous work optimizes compliance to better capture style [Liu et al. 2005], and measures forces to estimate compliance for resynthesis of motions involving contact [Kry and Pai 2006]. Compliance is also used to improve the robustness and naturalness of RL controllers. Lee et al. [2022] explore the strategy of providing a user-controlled parameter that modulates compliance, the inverse of stiffness, to produce more natural motion in the presence of external perturbations or variations in the task. Our work similarly demonstrates that RL control policies produce unnatural motions due to increased stiffness.

In addition to RL based approaches, other methods using PD control for motion synthesis may exhibit some of the problems we highlight in this paper. Liu et al. [2010] propose a sampling-based method for synthesizing physically valid motion in a contact-rich environment from motion capture clips. They further extend their method to generate longer motions from different clips by learning a feedback control to transition between clips [Liu et al. 2016]. Coros et al. [2010] demonstrate a motion control framework that generalizes well across different locomotion styles, where they use gravity compensation to avoid high PD gains. While many methods require motion capture data as

the reference, there are trajectory optimization methods that don't require motion capture or key poses, such as the work of Al Borno et al. [2012].

2.2 Human Motion Priors

Despite humanoid skeleton models having a high degree of articulation, low dimensionsal spaces have proven useful for representing and controlling human motion [Safonova et al. 2004]. Numerous work has addressed the problem of unnatural motion synthesis by learning natural human pose priors [Habibie et al. 2017; Ling et al. 2020; Yin et al. 2021]. Other approaches have enforced symmetry in walking gaits produced by controllers [Yu et al. 2018]. These approaches focus mainly on kinematics aspects of the motion. Our work instead considers characteristics of controllers that are important for ensuring that resulting motions are human-like.

The range of joint motions also plays an important role in producing realistic human motion. Several works have recognized that limits of joint motion are pose dependent. Akhter and Black [2015] learn a pose dependent model of joint limits using a dataset that explores the space of human articulation. Jiang and Liu [2018] learn an implicit function encoding joint limits and that can be included as part of the constraint solve in a physics simulation.

Jiang et al. [2019] demonstrate that more natural-looking motions can be synthesized by learning torque limits based on actuation controllers that are more complex and based on human muscle models. Findings in the biomechanics literature indicate that torque limits are dependent on both the joint angle and the rotational velocity at the joint [Anderson et al. 2007]. We take this into account in our analysis of how joint torque limits change, and we compare the torques produced by RL-based controllers to a data-driven model that we compute.

Other recent work has observed that imposing torque limits on the character model improves the naturalness of motion for some tasks. Tao et al. [2022] propose using a curriculum approach that gradually reduces character torques until it successfully performs the task with human-like limits.

2.3 Biomechanics

Human motion is characterized by variation. Recent studies of human gait have demonstrated that, even for seemingly straightforward tasks such as walking and running, foot placements are velocity dependent, in particular the step width and length [Collins and Kuo 2013; van der Zee et al. 2022]. Variation of motion and foot placement is further increased when there are perturbations, making human response to perturbation during locomotion challenging [Danforth et al. 2022]. Other work has focused on motor control adjustments to enrich variability [Scott 2002; Todorov and Jordan 2002].

Biomechanics literature has also examined the stiffness exhibited by humans when performing certain tasks. Bennett et al. [1992] found that arm joints are compliant when performing reaching tasks, and that low stiffness helps to accommodate deviations in the desired trajectory. Similar findings were obtained through experiments by Piovesan et al. [2013]. Our survey of papers on this topic found that the effective joint stiffness of the elbow, shoulder, knee, and ankle for real humans [Bennett et al. 1992; Gomi and Kawato 1997; Ludvig et al. 2012a; Piovesan et al. 2012, 2013] are much lower than typical values used for simulated humans in physics-based animation work [Bergamin et al. 2019; Peng et al. 2018a, 2021; Xu and Karamouzas 2021].

3 EFFECTIVE STIFFNESS OF A POLICY

Control policies often rely on low-level feedback controllers for joint actuation. Proportional derivative controllers are a popular choice since they are simple to use and perform well even at low query rates [Peng and van de Panne 2017]. The stiffness and damping parameters of these

controllers can be tuned, although they do not give a complete picture of the *effective stiffness* of the character. We define the effective stiffness as the torque response due to a change in the character's state, and compare its computed value with results reported in the biomechanics literature.

Consider the proportional derivative (PD) control law

$$\tau = -k_p(\theta - \bar{\theta}) - k_d(\dot{\theta} - \bar{\dot{\theta}}) \,. \tag{1}$$

Given a joint's position θ and velocity $\dot{\theta}$, the controller attempts to actuate the joint using control torque τ to realize the desired position and velocity, $\bar{\theta}$ and $\bar{\dot{\theta}}$ respectively.

An RL agent observes the virtual environment and performs an action according to the control policy given the current state, which includes the positions, orientations, and velocities of each body segment (for example, see [Peng et al. 2018a]). The target positions are then derived from the actions provided by the control policy; the target velocities are typically zero. Specifically, given a policy $\bar{\theta} = \pi(s, g)$ that outputs the target position according to the state s and task goal g, and combined with the PD law in Equation 1, the control torque generated from the policy is

$$\tau = -k_p(\theta - \pi(s, g)) - k_d\dot{\theta} \,. \tag{2}$$

The stiffness of the policy can therefore be defined and computed as

$$k_\pi \equiv -\frac{\partial \tau}{\partial \theta} \tag{3}$$

$$= k_p \left(1 - \frac{\partial \pi}{\partial \theta}\right) \tag{4}$$

$$= k_p \left(1 - \frac{\partial \pi}{\partial s} \frac{\partial s}{\partial \theta}\right), \tag{5}$$

where $\frac{\partial \pi}{\partial s}$ is computed from the policy, and $\frac{\partial s}{\partial \theta}$ is based on the current character state. Rather than computing the policy gradient with knowledge of the implementation (e.g, backpropagation through a neural network), we compute $\frac{\partial \pi}{\partial \theta}$ simply using finite differences with changes in the state vector with a step size of 1×10^{-8}.

One might ask, why not measure the stiffness of control policies in a manner resembling what is done for real humans? That is, we could apply perturbations and estimate the stiffness of different joints based on the measured response [Arami et al. 2020; Bennett et al. 1992]. However, this process is difficult and error prone in human experimentation and requires careful set-up, which has lead to limited examples in the real world. We initially devoted a good deal of effort to measuring stiffness in this way, but achieved minimal success obtaining consistent measurements. We speculate that this is due to the policy evaluation rate being slower than the simulation rate. Thus, θ in the PD law remains constant between policy evaluations, causing the torques to "spike" in the timestep immediately following a policy evaluation and exhibit an overdamped behavior afterward. This leads to a stiffness that depends on the simulator configuration. In the end, rather than struggling to produce results in a manner comparable to traditional human style measurements, we instead exploit the fact that we can directly measure quantities using the policy gradient, something that is otherwise impossible in a real-world experiment.

Using the expression in Equation 5, we evaluate the effective stiffness of RL controllers trained using several state-of-the-art methods. Figure 2 shows knee and elbow stiffness produced during walking and running motions. We observe that these controllers are significantly stiffer in comparison to the typical range used by humans for locomotion tasks as reported in the biomechanics literature by Ludvig et al. [2012b] and Piovesan et al. [2013].

There are likewise large non-smooth variations in the effective stiffness produced by the controllers, even for small changes in the state, in particular from the DM and AMP policies. Intuitively,
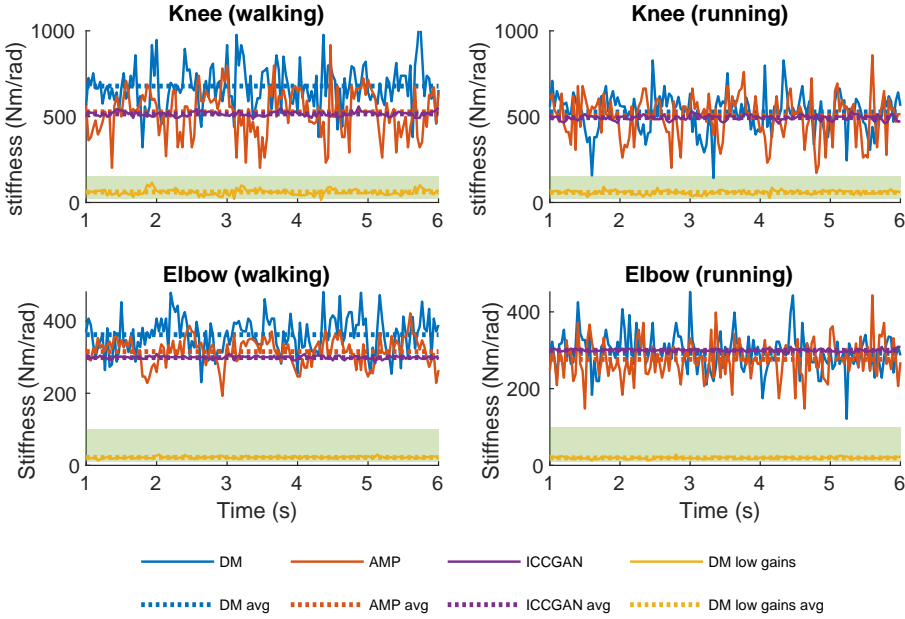
Fig. 2. Stiffness of the knee and elbow joints of walking (left) and running (right) control policies. The shaded green areas denote a plausible range of human joint stiffness reported in the biomechanics literature.

we anticipate that for a smoothly changing trajectory, the control policy should generate smooth stiffness changes. However, this is not the case, and our analysis indicates that the controller modulates the stiffness of each joint by as much as 200% from the nominal stiffness at the rate of the control updates. To our knowledge, no prior work has reported explicitly about stiffness modulation in humans. However, early work by Bennett et al. [1992] reported on time-varying stiffness of joints during voluntary upper body movement and their results do suggest smooth variations in the stiffness. We interpret the observed behavior as the controller compensating for the dynamics of a (possibly inadequately tuned) spring-damper servo in order to better track the desired trajectory.

To further explore this space, we retrained the DM walking and running controllers with lower gains. We reduce the knee stiffness from 500 Nm/rad in the original DM controller to 50 Nm/rad, and the elbow stiffness from 300 Nm/rad down to 30 Nm/rad. All the other parameters are the same as the original work. The effective stiffness of the retrained low-gain controllers also appears in Figure 2, and indeed the new curves lie within the reported natural range of human stiffnesses. But the trade-off is a larger imitation error and a harder training task that takes longer to converge. We revisit this with more discussion and data in Section 6.

4 JOINT TORQUE LIMITS

Although the range of human joint motion is directly measurable using motion capture equipment, measuring the torques that produce those motions is more challenging. In fact, angular PD servos do not exist in the human body, and instead actuation occurs due to muscular contractions. Work

in biomechanics has addressed this gap through careful experimental processes and indirect measurements that derive the effective torque produced during human motion. In this section, we compare the torques produced by control policies to ranges reported in the biomechanical literature. Furthermore, we estimate our own range of human torques via an inverse dynamics (ID) process applied to a large motion capture database. From the inverse dynamics data, we fit limit surfaces that give the pose and velocity-dependent torque limits for each joint such that the maximum joint torque $\tau_{\max}$ a joint can produce is a function of the joint position θ and joint velocity $\dot{\theta}$:

$$\tau_{\max} = F(\theta, \dot{\theta}). \tag{6}$$

4.1 Dataset Generation

To learn the function in Equation 6, we generate a dataset that is comprised of valid joint torques along with the joint position and velocity at each frame. We begin by extracting the position and velocity from a large motion capture database. An inverse dynamics algorithm is then used to compute joint torques $\mathbf{Q}$. The dynamical system can be stated as a system of second-order differential equations:

$$\mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}}) = \mathbf{Q} + \mathbf{J}^T\mathbf{f} \tag{7}$$

Here, $\mathbf{M}(\mathbf{q})$ is the generalized mass matrix, $\mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})$ is the Coriolis and centrifugal and gravity term, and $\mathbf{J}$ is the Jacobian matrix mapping joint velocities to spatial velocities at contact points. The mass properties of the character are computed using the mesh from the AMASS [Mahmood et al. 2019] database. We assume the density of the human body is 985 kg/m³. The velocities $\dot{\mathbf{q}}$ and accelerations $\ddot{\mathbf{q}}$ are approximated from sequences of joint positions using finite differences. Therefore, the goal of the inverse dynamics algorithm is to solve for the joint torques $\mathbf{Q}$ and contact forces $\mathbf{f}$ at each frame that satisfies the dynamical equations of motion.

We enforce for each contact c that the force $\mathbf{f}_c$ lies inside the Coulomb friction cone. The non-linear friction cone is approximated using a polygonal surface (i.e., a pyramid) and so a contact force $\mathbf{f}_c$ can be written as

$$\mathbf{f}_c = \mathbf{B}_c \boldsymbol{\lambda}_c, \tag{8}$$

where $\mathbf{B}_c$ is the basis matrix consisting of tangent and normal vectors, and $\boldsymbol{\lambda}_c = [\lambda_n, \lambda_{t_1}, \lambda_{t_2}, \lambda_{t_3}, \lambda_{t_4}]$ contains the magnitudes of the non-interpenetration force λ_n and frictional forces along each basis direction λ_{t_i}. Note that we have the additional constraints that

$$\boldsymbol{\lambda}_c \geq 0 \tag{9}$$

$$\left(\sum_{i=1}^{4} \lambda_{t_i}\right) \leq \mu\lambda_n, \tag{10}$$

where $\mu = 1$ is the friction coefficient used in our analyses.

Since humans cannot apply forces at their "root", we consider the physical model to be underactuated. Therefore, the first six components of the joint torque $\mathbf{Q}$ must be zero, since these correspond to the position and orientation of the root. Determining contact forces therefore becomes a constrained minimization problem:

$$\arg\min_{\boldsymbol{\lambda}} \left\| \mathbf{M}_{1:6,1:6}(\mathbf{q})\ddot{\mathbf{q}}_{1:6} + \mathbf{C}_{1:6}(\mathbf{q}, \mathbf{q}) - \mathbf{J}^T\mathbf{B}\boldsymbol{\lambda} \right\|.$$

The global matrix $\mathbf{B}$ and vector $\boldsymbol{\lambda}$ contain the contact basis and force magnitudes for all contacts. The above minimization problem can be rewritten as a quadratic programming problem, such that

$$\arg\min_{\mathbf{x}} \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x}, \tag{11}$$

where

$$\mathbf{x} = \boldsymbol{\lambda} \tag{12}$$

$$\mathbf{A} = \frac{1}{2} \mathbf{B}^T \mathbf{J} \mathbf{J}^T \mathbf{B} \tag{13}$$

$$\mathbf{b} = -\left(\mathbf{M}_{1:6,1:6}(\mathbf{q})\ddot{\mathbf{q}}_{1:6} + \mathbf{C}_{1:6}(\mathbf{q}, \dot{\mathbf{q}})\right)^T \mathbf{J}^T \mathbf{B} \tag{14}$$

and the inequalities in Equations 9 and 10 are satisfied. Joint torques $\mathbf{Q}$ are then recovered by direct computation by substituting the contact forces $\mathbf{f}$ into Equation 7.

The inverse dynamics algorithm is applied to sequences from the CMU and SFU motion capture databases. We use the AMASS [Mahmood et al. 2019] version of these databases to simplify the construction of the body model and tracking skeleton. Additionally, only foot and hand contact is detected for the inverse dynamics solve. We simply add contacts on the foot and hand body segments that have close to zero velocity and that are close to the ground plane. Although this might introduce some errors, there is no significant impact on our analysis since our goal is to estimate joint torque limits. After the algorithm is executed for all motion clips, we have the joint positions, velocities and torques for each frame in our training dataset.

4.2 Model Learning

To construct a torque limit model for each joint, we bin training samples according to their position and velocity. For single degree-of-freedom (DoF) joints, we store the maximum and minimum torque samples for a 29×24 grid of configurations across the range of joint positions and velocities. For the extension torque limit of each cell, the average of the top 10 percent of maximal torques is computed and used as the value of the cell (if there are fewer than 100 samples, we use the highest 10 samples). The same method is then used to generate the samples for the flexion limit, except that we compute the minimum instead of the maximum. For three DoF joints, we only consider the magnitudes of the joint position and joint velocity, rather than values along individual axes, to avoid the curse of dimensionality. This simplification assumes that the joint torque limits are uniform in all directions, which might not reflect a real human joint. However, this is a necessary compromise due to the size of the sample data.

After selecting minimum and maximum torques for each joint, we generate samples from the grid to train a neural network to fit the training data. Samples are randomly generated for different pose and velocity configurations and bilinear interpolation is used to compute values between grid cells, and 100 samples are generated per grid cell. We use $75,000$ samples for training the model. The neural network for the flexion limit uses 4 hidden layers with 10 nodes in each layer. The network for the extension limit uses 2 hidden layers with 10 nodes for the first layer and 5 nodes for the second layer. All the hidden layers use a tanh activation function.

Figure 4 is representative of the discrepancy between our analysis of human motion and that of the control policies generated. Controls routinely apply torques indicated in orange that exceed limits of our data-driven model (shown in blue) and the biological model proposed by Anderson et al. [2007] (shown in grey). Taken in isolation, each joint infrequently applies torques that are "out of bounds". However, in aggregation, these invalid torques have multiplicative effects that give the character transient super-human strength. This is most often revealed when the character is asked to perform tasks that deviate from the reference motion. This results not only in the character appearing too strong, but we hypothesize that this leads to policies finding solutions that are not realistic since human joints are particularly weak in some configurations. In other words, the muscles that power specific joints are not able to gain leverage or contraction in some joint configurations.
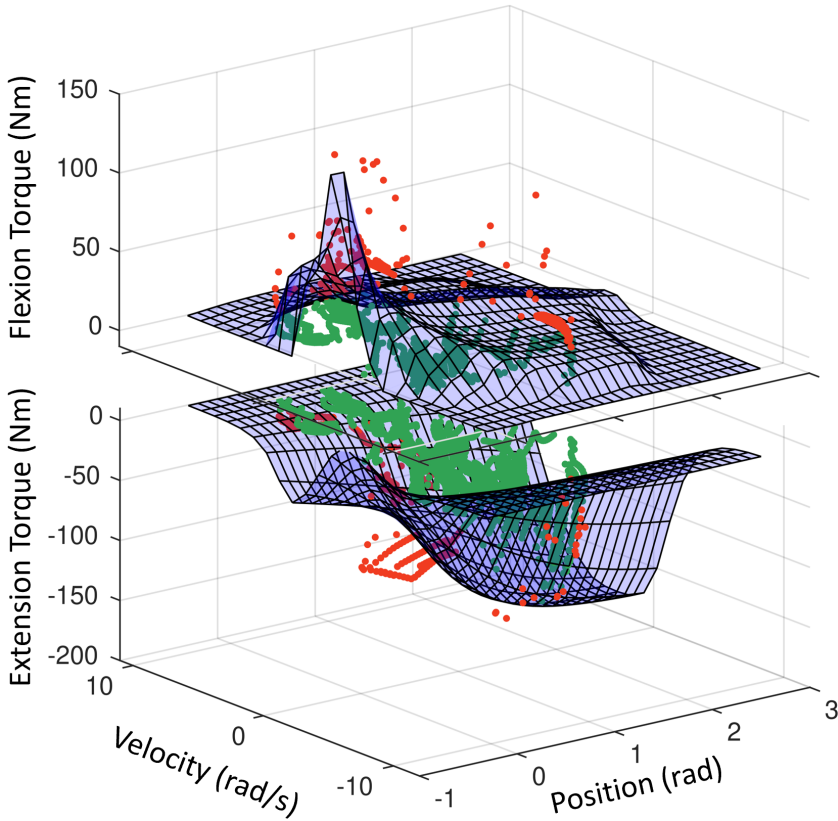
Fig. 3. Surfaces show the estimated knee torque limits learned from the sample data. Dots show the knee joint torques produced by a DeepMimic policy for a jumping motion. Green: torques inside our estimated joint torque limit. Red: Torques outside our estimated joint torque limit.

We expect that as more care is placed in shaping the torque limits, for example, to follow the manifolds shown in the figures, policies will exhibit more human-like strength, and fail in more coherent and reasonable fashions. From the figure, we can also see there are differences between the two models because of the nature of different sources. The biomechanics studies evaluate hard limits and provide a model for extreme human performance. Meanwhile, with ID applied to a wide set of examples, we identify torque limits from what was present in the data. While humans may be able to achieve the ideal limits, in practice they employ more conservative limits that are revealed in what we call the "expected" range.

Additionally, we observed that the policies generate "spiky" control torques, as can be seen in Figure 4. Work in the neuroscience community hypothesizes that large changes in the torque signal are due to high levels of stiffness that are required for rapid changes in motion [Gribble et al. 1998]. However, the control signals generated by the policies are far noisier and characterized by higher frequency than signals derived from neuroscience experiments (e.g., see Bennett [1993]). DeepMimic uses stable PD (SPD) [Tan et al. 2011] for low-level control, which is popular in many character animation frameworks due to its stability. We retrained a DeepMimic walking policy using PD control and a smaller simulation step size to ensure stability, and we observed similar
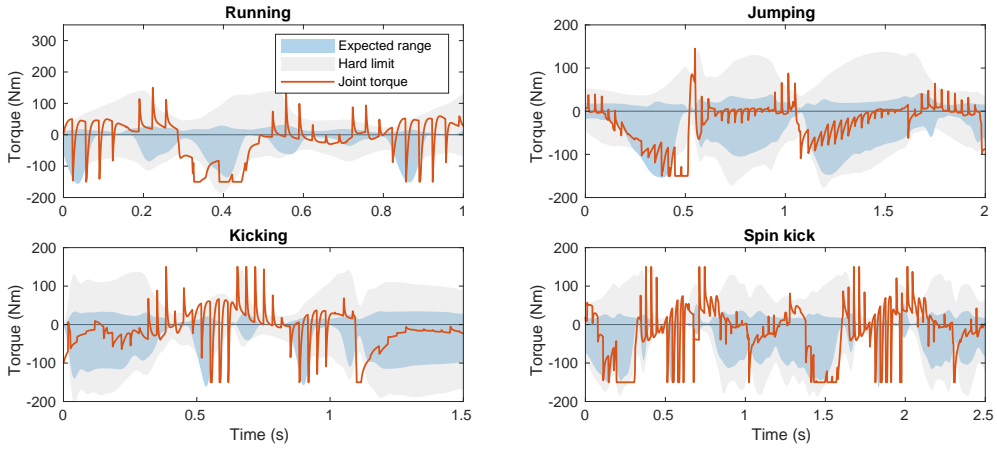
Fig. 4. Knee joint torques of Running, jumping, kicking and spin kick motion generated by DeepMimic, the expected torque range using our data-driven approach, and the estimated hard limit using the method proposed by Anderson et al. [2007].

"spiky" control torques. This suggests that PD-like control will generate unnatural torques when the policy evaluation rate is different from the simulation rate.

5   MOTION VARIABILITY

One way to test whether a control policy is too "smart" is to examine motion variability. We focus specifically on the variability of footsteps and hand-swinging motions. For footsteps, we compare the distribution of the lengths and widths between steps for real human motion to those in the motion generated by different control policies. For hand swinging, we make a comparison of controllers and recorded human motion in the frequency domain.

Our intuition is that humans are complicated systems, not controlled via precise and deterministic signals, and neither have perfect actuation. Athletes, dancers, and musicians practice frequently with the goal of reducing variability. Without training, it is either difficult or uncommon for humans to repeat the same trajectory exactly. Therefore, controllers for simulated humans should produce motion that contains a natural amount of variation, even without any perturbation from the environment. For our study of footsteps hand swing variability, we focus on undisturbed walking and running.

5.1   Foot Step Variability

The pattern of footsteps during locomotion is a topic of interest in biomechanics research. Collins and Kuo [2013] shows the footstep variability during walking along with the relationship between step length and walking speed. Inspired by this, we analyze and compare the footstep behaviors of real humans, recorded using motion capture equipment, with those observed in trajectories produced by state-of-the-art control policies.

We compare both footstep length and footstep width of walking motion. A footstep is detected when the velocity of the foot is below a threshold and close to the ground plane. The footstep length is defined as the distance between the current step location to that of the next, in the walking direction. The footstep width is defined similarly as the distance in the lateral (side) direction. We consider both footsteps from left to right and from right to left, and plot them separately because this reveals a preference for one foot over another (e.g., footedness).
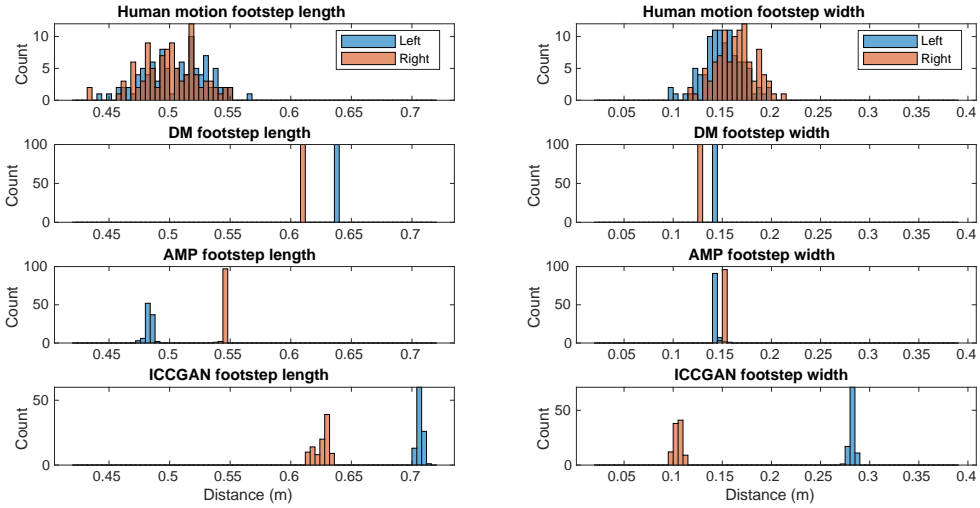
Fig. 5. The footstep length (left) and width (right) of walking motions from human motion data, DeepMimic, and AMP policies.
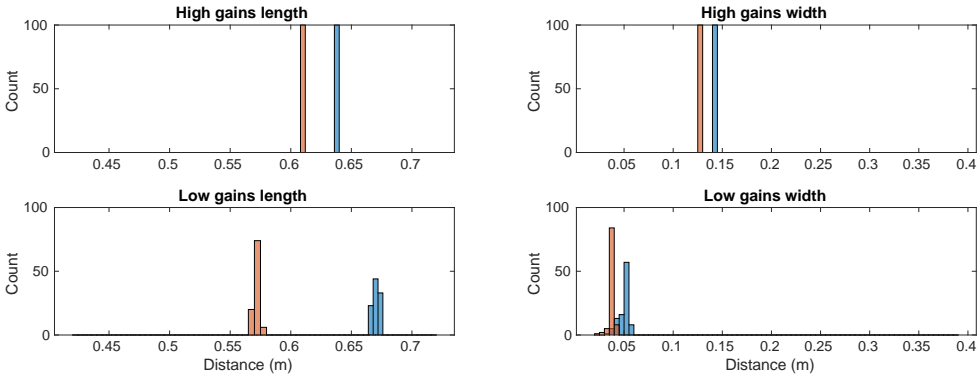

Fig. 6. The footstep length and width from a DeepMimic control policy with high (top) and low (bottom) PD gains.

Figure 5 shows the results. The human motion is derived from motion capture of a single subject on a treadmill set to a fixed speed to control for variations due to the correlation of step-length and speed. We also looked at shorter motion capture recordings in free environments and found similar results. The results show a stark difference between the human data and that of the physics characters. Namely, humans reveal a broad and well-distributed variation in their step lengths. In contrast to the step data produced by learned policies, which shows minimal variation, the human data shows that step lengths and widths vary over 10 cm or more, with no single value that is highly preferred. Humans are always varying their steps. Additionally, we observe that using lower PD gains slightly increases footstep variability, as shown in Figure 6. Finally, a surprising observation is that the walking policies exhibit consistently asymmetric stepping despite being trained to imitate a symmetric motion cycle.
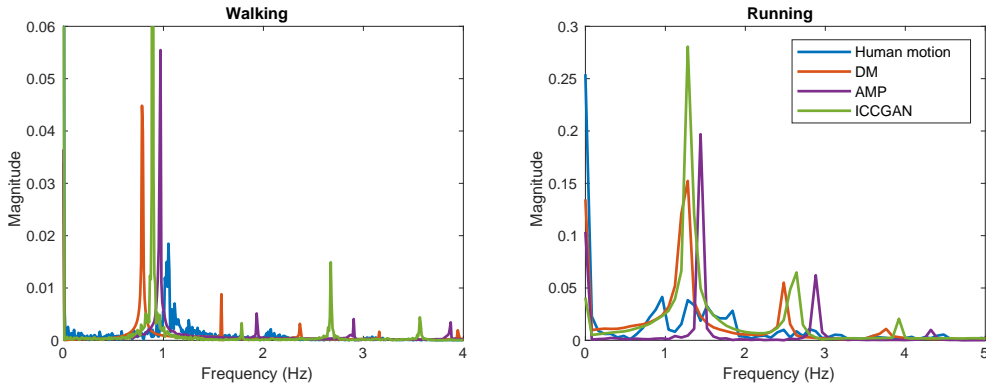
Fig. 7. The hand movement within the root body frame for walking (left) and running (right) from different policies and motion capture data, shown in the frequency domain.

5.2 Hand Swing Variability

Hand swing is useful to measure because the hand is at the end of the arm, and will reveal the presence of any motion variation throughout the body. To this end, we perform a Fourier transform on hand motions of human motion and trajectories produced by control policies, and compare them in the frequency domain. This choice is derived from the difficulty in their comparison in the time domain, since different humans (real and virtual) have different sizes which leads to different lengths of walking cycles and different swing magnitudes. In the frequency domain, when a motion has less variability, we observe narrow "spikes" representing the preferred periodic components in the motion. We naturally anticipate a wider variation in human motion and an indication that there are non-periodic components to the motion. Because we do not expect high frequency content in hand swings during locomotion, we focus only on frequencies lower than 5 Hz for this analysis.

What we see in Figure 7 is that all control policies that we evaluate exhibit extreme regularity around select preferred frequencies. This is seen in the plots as the spikes, noting that each policy has a significant primary and secondary frequency with limited deviation from these. We interpret this to mean that even in the hand position, which should show the greatest variation at the end of a free appendage, the resulting control is unnaturally precise. In contrast, the human hand motion expresses a broader range of frequencies, with significantly less preference for a single frequency. Humans use their arms as counterweights in locomotion to stabilize and respond to differences in locomotion, and we see this through the expression of many frequencies in the motion data. Comparatively, the controllers are extreme in their preference to specific frequencies. This may indicate that imitation is successful, but in the manner that is absent of the variations related to how humans use their arms to stabilize or counteract disturbances.

6   PERTURBATION AND EXPERIMENTAL CONTROL

The proposed metrics show repeated and severe discrepancies between what appear to be visually natural movements and those created by humans. While characters can imitate humans in appearance, the importance of these findings is tied to how the controllers act when they deviate from the specific tracks and reference motions. One key aspect of this is in interaction, for example, the response following an unexpected perturbation. With characters that are too stiff, too strong, and too smart, the reactions to disturbances will be unnatural, limiting how agents can be used in practice. And indeed, in many of the current findings, the evaluation of response to disturbance

is made based on robustness, most commonly the ability for a character to remain standing or walking amid a barrage of incoming projectiles. Thus as a measure of controller effectiveness, the characters should withstand a storm of boxes being thrown from all angles, even though this is highly unlike how a human may respond. Instead, we propose that the metrics shown here (among others) can help build controllers that are more natural in their response.

To support this view, we evaluate how controllers respond to perturbations in a controlled scenario where the hand of the character receives a modest (50 N) impulse for 0.2 s to see how the output is affected by the disturbance. Unfortunately, is it difficult to compare this to ground truth or biomechanical literature because such an experiment would require strict control, and repeatability, which is easily foiled by the presence of errors and unknowns – as we can infer from our variability analysis. Instead, to expose sensitivity to disturbances, we propose a set of purposeful changes to the control and verify that the result is what we expect.

Specifically, we lower the gain for the servo in the control, as mentioned in Section 3, to make the character's stiffness more humanlike. Putting the stiffness within human ranges alone under the same conditions, the character clearly exhibits a much more exaggerated (less stiff) response, which is quantitatively more natural (See Figure 2). We propose to couple this lower gain with a feedforward (FF) term that is derived from recording FF torques from an existing (high gain) policy. This is a common control technique from animation and robotics but has not appeared in the DRL literature to our knowledge. Such terms appeared in animation previously, such as [Yin et al. 2007]. Finally, we add a reaction delay as we posit that one way that policies can be too smart is that they can react instantaneously to an unexpected perturbation. By adding a purposeful delay in the system we deliberately increases the effect of the disturbance as the control is not able to react as quickly. Previous work in animation has accounted for such delays [Shiratori et al. 2009; Yeo et al. 2012; Zordan et al. 2005] but again this has not appeared in DRL research to our knowledge. We implement this set of modifications in sum as a new DM policy trained with: 1) low gains; 2) FF torques; and 3) a control delay which is built by the policy receiving a 100 ms old state vector.

Plots in Figure 8 show the impact of these three changes to the control compared to the unmodified response of DM and ICCGAN. The original controls' responses are nearly visually imperceivable as the characters *power through* the impulse, acting like kinematic characters in the presence of one-way collisions. The proposed modified controller, on the other hand, reveals a significantly more noticeable and quantitatively larger response as seen in the lower graph in Figure 7. We call the readers attention to the magnitude and delay of the return due to the response in contrast to the original policies. Notably, the hand is still deviating from the unperturbed motion nearly a second after the impulse (at about 1.3 s on the graph). We point out that ICCGAN is not phase locked to a given motion, and therefore does not return as explicitly as DM policies do.

To produce these results, we trained a new DM policy with the modifications specified: lower servo gains at one tenth of their original values; a built-in "sensor" delay where the controller of 100 ms in the past; and the described FF torques taken as a filtered average from the original (DM) policy and phase locked to the walk cycle. As a form of ablation, in comparison to the original policy, the impact of the lower gains and FF on training time was slight but noticeable, with no significant change in training performance after training. However, the reaction delay exhibited a notable performance reduction and took significantly more time to train. Further, the reaction delay failed to train when the FF term was removed. Empirically, we believe this is something that can be overcome, as the higher gain plus delay case did succeed, but further study is necessary to understand the nuance of the FF and delay's coupling. We also note sensor, neural, and muscle activation delay are present in humans [Stepan 2009] making the additional cost motivated as a means for improving naturalness.
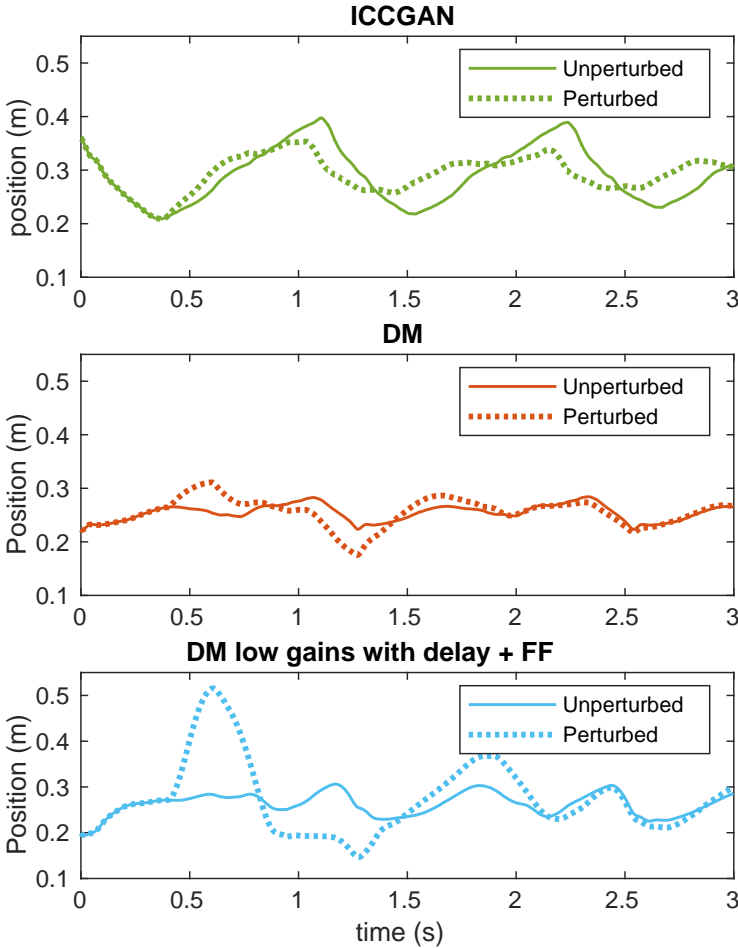
Fig. 8. Hand positions in character's local frame with and without a 50 N perturbation for 0.2 s starting at 0.4 s.

7 CONCLUSIONS AND FUTURE WORK

In this paper, we establish a set of metrics and baselines for evaluating control policies and employ these in the assessment of a number of recent approaches. Our evaluation of existing policies shows that although many controllers perform well in tracking human motions, there are still significant areas for improving the naturalness of control policies. For example, in order to generate torques that are large enough to actuate the character, especially for highly dynamic motions such as jumping, many methods use high PD-servo gains, which (we hypothesize) causes the motion to look stiffer than a human in the presence of a disturbance. One simple way to reduce the effective stiffness of the policy is using lower PD gains. A potentially better way to lower the effective stiffness is to control the gradient of the policy with respect to the joint states. This provides a more flexible way to control the effective stiffness, since the stiffness of the human joints usually changes in different phases. While time-varying stiffness is not discussed in this paper, we believe

it is a strong direction for future research and a means to make policies act more like humans in their stiffness.

Further, although many approaches limit joint torque levels, most set constant limits for the joints. This ignores the fact that the torque limit is a function of the joint position and velocity. It also causes an unnatural plateau in the torque profile. We reveal this plainly through a simple torque metric and establish important baseline comparisons with reported hard limits from biomechanics, and from a derived set of limits computed using inverse dynamics for data from a wide spectrum of human movements. The latter also offers one opportunity to have a more precise joint torque limit control, by using this computed value to limit joint torque. However, a limitation in this proposal is that our method only models torque limit as a function of angle and velocity while in reality, the joint torque limit is also affected by the state of the neighboring joints. This offers another promising path for future investigation but further research needs to be done to explore the modeling of this complicated relationship. In addition to violating joint torque limits, we also found that both PD control and SPD control show an unnatural "spiky" torque profile, which deviates from the smooth control signals suggested by the biomechanics literature [Hamner et al. 2010].

To evaluate the control's "intelligence", we propose to assess variability in this paper. And this metric indeed reveals some key differences between the control policies' outputs and human performance. In summary, our assessment shows that existing controllers produce behavior that is far too precise, meaning that they generate motions that lack the variability a real human exhibits. A clear potential fix is to inject more noise into the control policy, for example to mimic sensor error, variable delays, imperfect muscle actuation, and so on. However, we anticipate that these will also requires careful parameter tuning to match distributions of human behavior. Further, related literature suggests that humans need time to react to perturbations from the environment to remain natural which is not a behavior exhibited by the control policies we evaluated. Finding a quantitative way to measure the reflex and latency of physics-based characters is yet another direction for future work. But in sum, we feel our metrics for "smarts" are still limited and we hope to inspire others to propose new methods to assess the intelligence of control policies in contrast to humans. We also hope to spur others to adopt such evaluations in a means to move the domain past the use of solely visual or anecdotal evaluation methods to assess the success of DRL approaches.

REFERENCES

I. Akhter and M. J. Black. 2015. Pose-conditioned joint angle limits for 3D human pose reconstruction. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1446–1455. https://doi.org/10.1109/CVPR.2015.7298751

M. Al Borno, M. De Lasa, and A. Hertzmann. 2012. Trajectory optimization for full-body movements with complex contacts. *IEEE transactions on visualization and computer graphics* 19, 8 (2012), 1405–1414.

D. Anderson, M. Madigan, and M. Nussbaum. 2007. Maximum voluntary joint torque as a function of joint angle and angular velocity: Model development and application to the lower limb. *Journal of biomechanics* 40 (02 2007), 3105–13. https://doi.org/10.1016/j.jbiomech.2007.03.022

A. Arami, E. van Asseldonk, H. van der Kooij, and E. Burdet. 2020. A Clustering-Based Approach to Identify Joint Impedance During Walking. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28, 8 (2020), 1808–1816. https://doi.org/10.1109/TNSRE.2020.3005389

D. Bennett, J. Hollerbach, and Y. Xu. 1992. Time-varying stiffness of human elbow joint during cyclic voluntary movement. *Experimental Brain Research* 88, 2 (1992).

D. J. Bennett. 1993. Torques generated at the human elbow joint in response to constant position errors imposed during voluntary movements. *Experimental Brain Research* 95, 3 (1993), 488–498.

K. Bergamin, S. Clavet, D. Holden, and J. R. Forbes. 2019. DReCon: Data-Driven Responsive Control of Physics-Based Characters. *ACM Trans. Graph.* 38, 6, Article 206 (nov 2019), 11 pages. https://doi.org/10.1145/3355089.3356536

S. H. Collins and A. D. Kuo. 2013. Two Independent Contributions to Step Variability during Over-Ground Human Walking. *PLOS ONE* 8, 8 (08 2013), null. https://doi.org/10.1371/journal.pone.0073597

S. Coros, P. Beaudoin, and M. Van de Panne. 2010. Generalized biped walking control. *ACM Transactions On Graphics (TOG)* 29, 4 (2010), 1–9.

S. M. Danforth, X. Liu, M. J. Ward, P. D. Holmes, and R. Vasudevan. 2022. Predicting Sagittal-Plane Swing Hip Kinematics in Response to Trips. In *2022 9th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*. 1–8. https://doi.org/10.1109/BioRob52689.2022.9925354

H. Gomi and M. Kawato. 1997. Human arm stiffness and equilibrium-point trajectory during multi-joint movement. *Biological Cybernetics* 76, 3 (1997), 163–171.

P. L. Gribble, D. J. Ostry, V. Sanguineti, and R. Laboissière. 1998. Are complex control signals required for human arm movement? *Journal of neurophysiology* 79, 3 (1998), 1409–1424.

I. Habibie, D. Holden, J. Schwarz, J. Yearsley, and T. Komura. 2017. A recurrent variational autoencoder for human motion synthesis. In *28th British Machine Vision Conference*.

S. R. Hamner, A. Seth, and S. L. Delp. 2010. Muscle contributions to propulsion and support during running. *Journal of biomechanics* 43, 14 (2010), 2709–2716.

Y. Jiang and C. K. Liu. 2018. Data-Driven Approach to Simulating Realistic Human Joint Constraints. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*. 1098–1103. https://doi.org/10.1109/ICRA.2018.8461010

Y. Jiang, T. Van Wouwe, F. De Groote, and C. K. Liu. 2019. Synthesis of Biologically Realistic Human Motion Using Joint Torque Actuation. *ACM Trans. Graph.* 38, 4, Article 72 (jul 2019), 12 pages. https://doi.org/10.1145/3306346.3322966

P. G. Kry and D. K. Pai. 2006. Interaction Capture and Synthesis. *ACM Trans. Graph.* 25, 3 (jul 2006), 872–880. https://doi.org/10.1145/1141911.1141969

A. Kwiatkowski, E. Alvarado, V. Kalogeiton, C. K. Liu, J. Pettré, M. van de Panne, and M.-P. Cani. 2022. A Survey on Reinforcement Learning Methods in Character Animation. *Computer Graphics Forum* 41, 2 (2022), 613–639. https://doi.org/10.1111/cgf.14504 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.14504

S. Lee, P. S. Chang, and J. Lee. 2022. Deep Compliant Control. In *ACM SIGGRAPH 2022 Conference Proceedings* (Vancouver, BC, Canada) *(SIGGRAPH '22)*. Association for Computing Machinery, New York, NY, USA, Article 23, 9 pages. https://doi.org/10.1145/3528233.3530719

H. Y. Ling, F. Zinno, G. Cheng, and M. Van De Panne. 2020. Character Controllers Using Motion VAEs. *ACM Trans. Graph.* 39, 4, Article 40 (aug 2020), 12 pages. https://doi.org/10.1145/3386569.3392422

C. K. Liu, A. Hertzmann, and Z. Popović. 2005. Learning Physics-Based Motion Style with Nonlinear Inverse Optimization. *ACM Trans. Graph.* 24, 3 (jul 2005), 1071–1081. https://doi.org/10.1145/1073204.1073314

L. Liu, M. V. D. Panne, and K. Yin. 2016. Guided learning of control graphs for physics-based characters. *ACM Transactions on Graphics (TOG)* 35, 3 (2016), 1–14.

L. Liu, K. Yin, M. Van de Panne, T. Shao, and W. Xu. 2010. Sampling-based contact-rich motion control. In *ACM SIGGRAPH 2010 papers*. 1–10.

D. Ludvig, S. Pfeifer, X. Hu, and E. J. Perreault. 2012a. Time-Varying System Identification for Understanding the Control of Human Knee Impedance. *IFAC Proceedings Volumes* 45, 16 (2012), 1306–1310. https://doi.org/10.3182/20120711-3-BE-2027.00410 16th IFAC Symposium on System Identification.

D. Ludvig, S. Pfeifer, X. Hu, and E. J. Perreault. 2012b. Time-Varying System Identification for Understanding the Control of Human Knee Impedance. *IFAC Proceedings Volumes* 45, 16 (2012), 1306–1310. https://doi.org/10.3182/20120711-3-BE-2027.00410 16th IFAC Symposium on System Identification.

N. Mahmood, N. Ghorbani, N. F. Troje, G. Pons-Moll, and M. Black. 2019. AMASS: Archive of Motion Capture As Surface Shapes. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE Computer Society, Los Alamitos, CA, USA, 5441–5450. https://doi.org/10.1109/ICCV.2019.00554

L. Mourot, L. Hoyet, F. Le Clerc, F. Schnitzler, and P. Hellier. 2022. A Survey on Deep Learning for Skeleton-Based Human Animation. *Computer Graphics Forum* 41, 1 (2022), 122–157. https://doi.org/10.1111/cgf.14426 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.14426

H. Park, R. Yu, Y. Lee, K. Lee, and J. Lee. 2022. Understanding the stability of deep control policies for biped locomotion. *The Visual Computer* (2022), 1–15.

X. B. Peng, P. Abbeel, S. Levine, and M. van de Panne. 2018a. DeepMimic: Example-Guided Deep Reinforcement Learning of Physics-Based Character Skills. *ACM Trans. Graph.* 37, 4, Article 143 (jul 2018), 14 pages. https://doi.org/10.1145/3197517.3201311

X. B. Peng, A. Kanazawa, J. Malik, P. Abbeel, and S. Levine. 2018b. SFV: Reinforcement Learning of Physical Skills from Videos. *ACM Trans. Graph.* 37, 6, Article 178 (dec 2018), 14 pages. https://doi.org/10.1145/3272127.3275014

X. B. Peng, Z. Ma, P. Abbeel, S. Levine, and A. Kanazawa. 2021. AMP: Adversarial Motion Priors for Stylized Physics-Based Character Control. *ACM Trans. Graph.* 40, 4, Article 144 (jul 2021), 20 pages. https://doi.org/10.1145/3450626.3459670

X. B. Peng and M. van de Panne. 2017. Learning Locomotion Skills Using DeepRL: Does the Choice of Action Space Matter?. In *Proceedings of the ACM SIGGRAPH / Eurographics Symposium on Computer Animation* (Los Angeles, California) *(SCA '17)*. Association for Computing Machinery, New York, NY, USA, Article 12, 13 pages. https://doi.org/10.1145/3099564.3099567

D. Piovesan, A. Pierobon, P. Dizio, and J. Lackner. 2012. Measuring Multi-Joint Stiffness during Single Movements: Numerical Validation of a Novel Time-Frequency Approach. *PLoS ONE* 7, 3 (08 2012). https://doi.org/10.1371/journal.pone.0033086

D. Piovesan, A. Pierobon, P. Dizio, and J. Lackner. 2013. Experimental Measure of Arm Stiffness During Single Reaching Movements with a Time-Frequency Analysis. *Journal of neurophysiology* 110, 10 (08 2013). https://doi.org/10.1152/jn.01013.2012

A. Safonova, J. K. Hodgins, and N. S. Pollard. 2004. Synthesizing Physically Realistic Human Motion in Low-Dimensional, Behavior-Specific Spaces. In *ACM SIGGRAPH 2004 Papers* (Los Angeles, California) *(SIGGRAPH '04)*. Association for Computing Machinery, New York, NY, USA, 514–521. https://doi.org/10.1145/1186562.1015754

S. H. Scott. 2002. Optimal strategies for movement: success with variability. *nature neuroscience* 5, 11 (2002), 1110–1111.

T. Shiratori, B. Coley, R. Cham, and J. K. Hodgins. 2009. Simulating Balance Recovery Responses to Trips Based on Biomechanical Principles. In *Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (New Orleans, Louisiana) *(SCA '09)*. Association for Computing Machinery, New York, NY, USA, 37–46. https://doi.org/10.1145/1599470.1599475

G. Stepan. 2009. Delay effects in the human sensory system during balancing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367, 1891 (2009), 1195–1212.

J. Tan, K. Liu, and G. Turk. 2011. Stable Proportional-Derivative Controllers. *IEEE Comput. Graph. Appl.* 31, 4 (jul 2011), 34–44. https://doi.org/10.1109/MCG.2011.30

T. Tao, M. Wilson, R. Gou, and M. van de Panne. 2022. Learning to Get Up. In *ACM SIGGRAPH 2022 Conference Proceedings* (Vancouver, BC, Canada) *(SIGGRAPH '22)*. Association for Computing Machinery, New York, NY, USA, Article 47, 10 pages. https://doi.org/10.1145/3528233.3530697

E. Todorov and M. I. Jordan. 2002. Optimal feedback control as a theory of motor coordination. *Nature neuroscience* 5, 11 (2002), 1226–1235.

T. J. van der Zee, E. M. Mundinger, and A. D. Kuo. 2022. A biomechanics dataset of healthy human walking at various speeds, step lengths and step widths. *Scientific Data* 9, 1 (2022).

P. Xu and I. Karamouzas. 2021. A GAN-Like Approach for Physics-Based Imitation Learning and Interactive Character Control. *Proc. ACM Comput. Graph. Interact. Tech.* 4, 3, Article 44 (sep 2021), 22 pages. https://doi.org/10.1145/3480148

S. H. Yeo, M. Lesmana, D. R. Neog, and D. K. Pai. 2012. Eyecatch: Simulating Visuomotor Coordination for Object Interception. *ACM Trans. Graph.* 31, 4, Article 42 (jul 2012), 10 pages. https://doi.org/10.1145/2185520.2185538

K. Yin, K. Loken, and M. Van de Panne. 2007. Simbicon: Simple biped locomotion control. *ACM Transactions on Graphics (TOG)* 26, 3 (2007), 105–es.

Z. Yin, Z. Yang, M. Van De Panne, and K. Yin. 2021. Discovering Diverse Athletic Jumping Strategies. *ACM Trans. Graph.* 40, 4, Article 91 (jul 2021), 17 pages. https://doi.org/10.1145/3450626.3459817

W. Yu, G. Turk, and C. K. Liu. 2018. Learning Symmetric and Low-Energy Locomotion. *ACM Trans. Graph.* 37, 4, Article 144 (jul 2018), 12 pages. https://doi.org/10.1145/3197517.3201397

V. B. Zordan, A. Majkowska, B. Chiu, and M. Fast. 2005. Dynamic response for motion capture animation. *ACM Transactions on Graphics (TOG)* 24, 3 (2005), 697–701.