Personalized Dance Synthesis Based on Physical and Cognitive Intensities

Xulong Tang*
The University of Texas at Dallas

Eun Yeo
The University of Texas at Dallas

Ruiyu Mao
The University of Texas at Dallas

Xiaohu Guo
The University of Texas at Dallas

Rawan Alghofaili
The University of Texas at Dallas

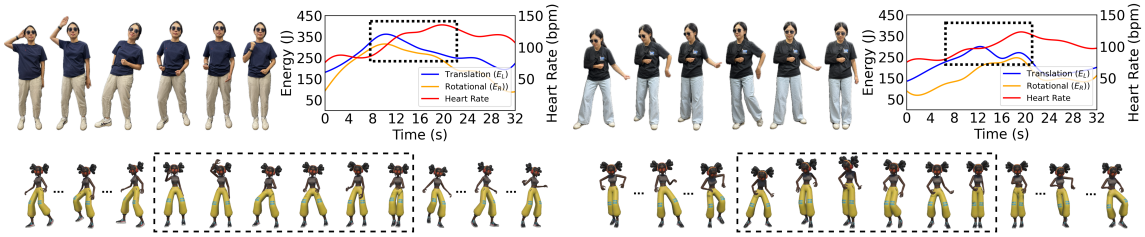


Figure 1: Our framework synthesizes dance routines with controllable physical and cognitive intensities by combining multiple cost terms. By constraining the intensity of specific joint groups (Right), the system can produce fixed-arm dance routines while maintaining the same physical intensity as an unconstrained routine (Left).

Abstract

Dance-based exergames like Just Dance can be a fun way to boost your fitness and sharpen your mind. However, designing the dance routines requires expertise in modeling and animation. We introduce an augmented reality (AR) personalized dance generation framework that synthesizes dance routines according to specified physical and cognitive intensities. Our system utilizes a curated library of motion-capture dance segments, which are intelligently combined through an optimization process to meet user-defined intensity and cognitive goals. This optimization also ensures smooth transitions between movements for natural dance flow. Users can customize routines by specifying physical constraints or injuries. Implemented in a depth-camera-based exergame that provides real-time performance feedback, our framework was evaluated through experiments and user studies confirming its effectiveness in generating personalized routines with varying levels of physical and cognitive intensity.

Index Terms: Exergaming, Generative Design, Optimization, Augmented Reality.

1 Introduction

Dance-based exergames combine social interaction with physical activity, driving their commercial success to new heights. Popular titles such as Just Dance [53], Dance Central [20], and Zumba Fitness [38] have attracted millions of active users worldwide, highlighting their widespread appeal and effectiveness in engaging diverse audiences in physical exercise [12].

Research has demonstrated that dance improves both physical fitness and cognitive function. Physically, dance involves sustained aerobic exertion, enhancing cardiovascular endurance, muscle strength, and balance [21]. Cognitively, dance engages neural processes related to memory, spatial awareness, and sensorimotor coordination, improving cognitive performance and brain health [45].

*e-mail: Xulong.Tang@UTDallas.edu

Advances in affordable motion-tracking technology, such as the Femto Bolt depth camera, have facilitated research into dance exercise games. Previous research using similar technologies has explored performance evaluation [1], physical rehabilitation [50], exertion-based games [39], and interactive choreography [8, 19]. Despite this progress, existing systems typically neglect the explicit control of cognitive and physical intensity. When intensity control is included, the resulting movements are limited to simplistic poses lacking the complexity and variety characteristic of real dance movements [66].

Effective dance training systems should produce physically and mentally engaging routines that users can easily remember. Recent research highlights that memorability in dance sequences improves when movements involve smaller, repetitive patterns [45]. Additionally, Basak et al. [4,5] found that tasks requiring cognitive control and memory benefit significantly from exercises that combine physical engagement with active cognitive involvement. This finding underscores the need for dance training systems that produce routines that are not only physically and cognitively challenging but also memorable.

To address these limitations, we propose an augmented reality (AR) dance synthesis framework that generates dance routines optimized for physical intensity and cognitive training. Our framework relies on a library of curated dance segments extracted from motion-capture dance datasets. After the user sets their training goals by specifying the routine's desired physical intensity and memorability, we sample and combine these dance segments during our optimization to generate the routines. The framework also matches the dance routine's length to the duration set by the user. To ensure our framework generates smooth and realistic dance routines, we included a transition cost that penalizes sudden changes between dance segments in our optimization.

Overall, the major contributions of our work include:

- We introduce an optimization-based framework that generates personalized dance routines by controlling translational and rotational kinetic energy to match target physical intensity, and by incorporating motion variation and duration constraints to achieve cognitive difficulty.
- We evaluated the effectiveness of our framework in generating physical and cognitive challenging dance routines using experiments and a user study.

2642-5254/26/$31.00 ©2026 IEEE
DOI 10.1109/VR67842.2026.00048

2 Related Work

Our work combines exergaming, AR, and procedural content generation to enable intensity-aware adaptive dance training. We analyze prior research to highlight gaps addressed by our contributions.

2.1 Dance Motion Generation

Recent years have witnessed significant progress in dance motion generation, where the goal is to synthesize realistic full-body human dance sequences. Early approaches typically employed encoder–decoder architectures to directly map input conditions to entire motion sequences [9, 29, 30, 59]. In the context of generative modeling, researchers have also applied Generative Adversarial Networks (GANs) to enhance motion realism by training discriminators to distinguish between real and generated dance sequences, thereby guiding generators toward producing more naturalistic outputs [8, 61]. More recently, Diffusion Models have been adopted for motion synthesis, achieving strong performance in generation [28, 31–33]. In parallel, efforts have been made to introduce control into dance generation, such as Disco [54], which enables control over style and rhythm factors, Keyframe Control [64] which uses keyframe interpolation to enforce beat-aligned, smooth motion. While these deep learning approaches produce visually compelling results, they lack mechanisms for controlling the physical intensity levels within the generated dance routines. Additionally, do they allow for adjustments to the variety and repetition of dance movements, which are essential for controlling the memorability of the choreography. Both of these factors are crucial in the context of physical and cognitive therapy [6]. By contrast, optimization-based [48, 63] approaches enable explicit control over translational and rotational intensity as well as motion variety. This user-centered design allows the system to generate dance routines that align with individual needs for different intensity and variety levels, supporting personalized training and rehabilitation.

2.2 Exergaming in Physical and Cognitive Therapy

Commercial dance exergames such as Just Dance adopt fixed choreography designed by professionals, which restricts adaptability across users with varying physical abilities and cognitive needs. This fixed-level design often fails to offer appropriate physical and cognitive challenges, limiting personalization and training efficacy.

Most prior research in exertion games emphasizes cardiovascular metrics such as heart rate [39]. However, heart rate alone does not fully capture the biomechanical demands of dance. Accurate dance performance requires additional kinematic indicators, including joint positions and rotations, which better reflect motor precision and physical exertion.

Exergaming and mixed reality dance therapies have shown promise in improving both physical movement quality and cognitive function across different populations. Ogawa et al. [44] conducted a systematic review demonstrating that exergaming in older adults can enhance cognition and dual-task performance, with potential to reduce fall risk. Similarly, Liu et al. [36] proposed a mixed reality dance movement therapy for children with ASD, using a virtual twin agent to guide the child, and observed improvements in movement quality and responsiveness to target tasks under varied training conditions.

Clinical exergaming systems typically rely on therapists to manually adjust intensity [51]. Such approaches are labor-intensive and not scalable for home-based or autonomous use. Recent studies highlight the importance of automated, individualized intensity calibration using real-time motion analysis [50]. Our framework automatically generates dance sequences tailored to diverse physical and cognitive training goals.

2.3 AR/VR Physical Training

AR/VR-based physical training platforms often lack fine-grained personalization mechanisms. For example, Sarupuri et al. [47] showed users benefit from personalized dance-based VR training, but most existing systems predefine difficulty levels without dynamic adaptation [7, 42]. AR choreography tools such as ChoreoMaster [8] and ChoreoCraft [19] focus on authoring support and creativity, rather than physical training outcomes. In contrast, our system generates diverse and challenging dance routines by modeling physical effort through biomechanical features.

Procedural approaches have been used to adjust exercise intensity in exergames. Li et al. [34] proposed optimizing cycling paths using estimated energy expenditure. However, such approaches are less suited for dance, which requires metrics beyond calorie estimates. Zhang et al. [66] explored optimizing pose-based motion games using joint rotation and center-of-mass metrics. While effective for isolated poses, their method was limited to static pose sequences and small motion variations. Our method extends this idea to dynamic dance sequences. Beyond exergaming, the broader graphics and animation community has investigated motion synthesis through optimization and probabilistic sampling. Kim et al. [25] introduced tiling of motion patches to generate continuous human motion, while MCMC-based methods [48, 63] demonstrated how sampling with constraints can support scalable and realistic scene generation. By optimizing over kinetic energy and full-body motion patterns, our system enables more realistic and physically engaging training experiences.

2.4 AR/VR Cognitive Training

AR/VR platforms have been explored for cognitive training, offering immersive environments that can enhance various cognitive functions. Virtual reality (VR) interventions have demonstrated moderate positive effects on overall cognition, attention/executive function, memory, and global cognition in older adults with mild cognitive impairment or dementia [67]. Similarly, augmented reality (AR) applications, such as the MarketMind AR system—a supermarket-themed serious game—have been utilized to train memory, attention, and executive function using mobile phone sensors. Studies have reported that older adults find such AR-based cognitive training tools usable and engaging, with potential cognitive benefits [23].

Our system introduces a controllable diversity mechanism that adjusts dance sequence variability. This allows the design of routines with different cognitive demands, facilitating structured cognitive training through movement.

3 Overview

Figure 2 illustrates the core framework of our system. We begin by randomly initializing a dance using dance segments sampled from a curated dance library. This initial sequence is then iteratively refined through an optimization procedure. At each iteration, the system evaluates candidate moves based on a weighted combination of cost terms, including duration, transition smoothness, motion variety, and physical intensity. The optimization process continues until a threshold is reached.

4 Creating the Dance Library

4.1 Data

To support our optimization process, we constructed a comprehensive library of dance motion data. This library was built by extracting and processing motion data from several publicly available datasets: AIST++ [30], AIOZ-GDANCE [29], FineDance [33]. For datasets featuring multiple dancers, we decomposed scenes into individual dance tracks, treating each dancer's performance as an independent sequence. After removing unusable or noisy

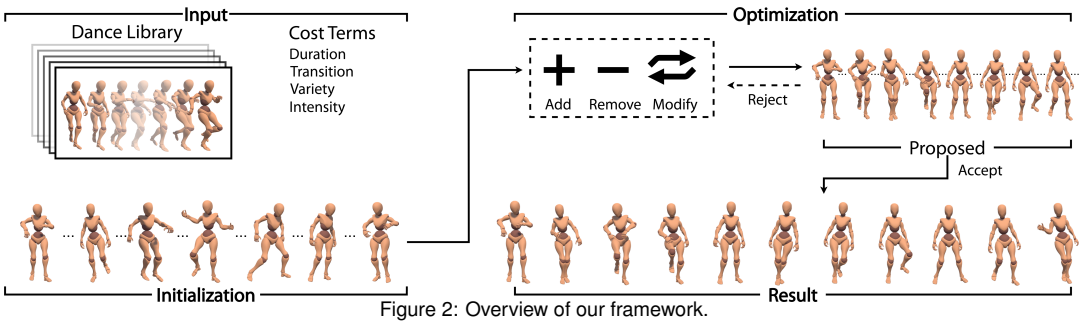


Figure 2: Overview of our framework.

sequences through our data cleaning pipeline, the final integrated dataset comprises approximately 38.5 hours of single-dancer motion data across 25 distinct genres.

The aforementioned datasets contain sequences of dance poses in varying formats, noise, and scale. As a result, we implemented a pipeline for preprocessing (Section 4.2), and segmentation (Section 4.3) to create our final dance library.

4.2 Preprocessing

First, we converted the extracted data into a unified representation to facilitate the sampling process during the optimization. The original datasets adopted different human body models, coordinate systems, and frame rates.

Pose Representation. To simplify our dane generation, we converted all motion data to the Skinned Multi-Person Linear (SMPL) model [37] representation and applied a series of spatial normalization steps. We then applied the SMPL forward kinematics to compute the 3D position of 24 key body joints. We also centered all poses around the pelvis joint and translated them to the origin $(0, 0, 0)$ in the global coordinate system. To ensure consistency, we rotated all sequences to face a unified forward direction (Z-forward) in a right-handed coordinate system. Finally, we normalized the joint positions using Z-score normalization per joint dimension across the dataset.

Noise Reduction. The motion capture datasets included some abnormalities that we addressed using our noise reduction pipeline. First, we iterated over every frame in each dance and calculated the acceleration of each joint. Any change in acceleration exceeding the joint threshold of two standard deviations above the average acceleration of that joint across the dance was detected as abrupt. These abrupt changes often indicated unnatural movements, typically caused by errors in the motion capture data. We removed the flagged frame and retained the preceding frames in our dataset if they were longer than 4 seconds.

Next, we manually reviewed these segments using Blender for visualization. We verified the presence of anomalies, such as twisted limbs and impossible joint rotations. We then removed these problematic frames and preserved only the valid segments.

Re-Sampling. We re-sampled all motion data to a fixed frame rate of 30 FPS to ensure temporal consistency across datasets.

4.3 Segmentation

After preprocessing, we segmented each dance sequence into fixed-length clips of 4 seconds. This segmentation strategy aligns with common musical structures—at a typical tempo of 120 BPM, a 4-second interval corresponds to exactly 8 beats, which often forms a complete movement phrase in choreographic practice [43, 62]. Prior works have adopted similar durations to facilitate alignment between musical rhythm and dance motion, supporting synchronization and stylistic coherence.

Segments shorter than 4 seconds were discarded. In total, we extracted approximately 34,613 valid segments from 38.5 hours of cleaned motion data. Each segment of 4 seconds in our library is composed of 120 poses because we re-sampled all motion data at 30 FPS. These segments constitute our *Dance Library*, which serves as the foundation for generating choreography optimized for intensity, smoothness, variety, and duration (Section 5).

5 Dance Generation

The goal of our system is to synthesize a dance composed of multiple segments, achieving a specified level of physical intensity, variety, and total duration. We also consider how smoothly transitions occur between segments to generate more realistic and fluid dances. These considerations are encoded as cost terms in our optimization.

5.1 Problem Formulation

We define a proposed dance $\mathcal{S} = (s_1, \ldots, s_n)$ as a sequence of dance segments s_i where each segment is sampled from our dance library $s_i \in \mathcal{D}$. Ultimately, each segment is composed of a sequence of skeletal poses that can be represented in $\mathcal{S}$ as $(p_1, \ldots, p_m)$, where m is the total number of poses.

Using the SMPL model, each pose within a segment contains the position and rotation of 24 articulated body joints $\mathcal{J} = \{j_1, j_2, \ldots, j_{24}\}$. We represent the position of joint j in pose p with a 3D vector $\mathbf{o}_{p,j} \in \mathbb{R}^3$. We represent the rotation of joint j within the same pose with respect to its parent using a rotation matrix $\mathbf{R}_{p,j} \in SO(3)$.

Total Cost. The objective of our framework is to synthesize a dance with smooth transitions between segments that meets a target intensity, variety, and duration. We evaluate a proposed solution $\mathcal{S}$ using the following total cost function:

$$\begin{aligned} C_{\text{Total}}(\mathcal{S}) = &\lambda_{\text{D}} C_{\text{D}}(\mathcal{S}) + \lambda_{\text{T}} C_{\text{T}}(\mathcal{S}) + \lambda_{\text{V}} C_{\text{V}}(\mathcal{S}) \\ &+ \lambda_{\text{LIN}} C_{\text{LIN}}(\mathcal{S}) + \lambda_{\text{ROT}} C_{\text{ROT}}(\mathcal{S}), \end{aligned} \tag{1}$$

where $\mathcal{S} = (s_1, \ldots, s_n)$ is a dance consisting of a sequence of segments from our dance library $\mathcal{D}$. Our total cost function is composed of five terms: a duration cost $C_{\text{D}}(\mathcal{S})$ to control the amount of time it takes to perform the dance, a transition cost $C_{\text{T}}(\mathcal{S})$ to encourage seamless transitions between consecutive dance poses, a variety cost $C_{\text{V}}(\mathcal{S})$ that controls the diversity of the segments selected in the optimization, and intensity costs $C_{\text{LIN}}(\mathcal{S})$ and $C_{\text{ROT}}(\mathcal{S})$ that set the desired level of physical exertion. We use the weight parameters $\lambda_{\text{D}}, \lambda_{\text{T}}, \lambda_{\text{V}}, \lambda_{\text{LIN}}$, and λ_{ROT} to control the influence of each cost term on the total cost. Users can adjust these parameters to prioritize different optimization goals, such as generating a more physically demanding dance or promoting a higher variety of segments in the dance.

5.2 Duration Cost

The duration cost ensures the synthesized dance sequence matches the target duration ρ_{D}. We penalize deviations from this target using a Gaussian cost to smoothly reduce the cost as the dance duration approaches the target duration. This cost allows us to directly control the length of the dance. Because longer dances tend to be more difficult to memorize, this also indirectly influences cognitive training. We define the duration cost of the dance as:

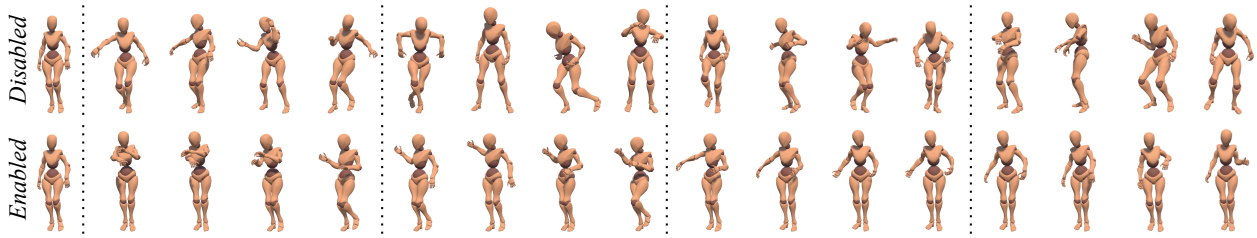


Figure 3: Removing the transition cost (top) results in abrupt transitions between segments, leading to an overall less fluid dance routine. The same duration, intensity, and variety costs were used to generate both dances.

$$C_{\mathrm{D}}(\mathcal{S}) = 1 - \exp\left(-\frac{\left(\sum_s t(s) - \rho^{\mathrm{D}}\right)^2}{2\sigma_{\mathrm{D}}^2}\right) \tag{2}$$

where $s \in \mathcal{S}$ refers to a dance segment in the proposed dance $\mathcal{S}$ while $t(s)$ computes the duration of the segment in seconds. ρ^{D} is the target duration of the sequences in seconds, σ_{D} controls the tolerance for deviations from ρ^{D}. A lower value of C_{D} indicates that the total duration of the generated dance is closer to the target ρ_{D}.

5.3 Transition Cost

Ensuring continuity across stitched motion segments has been extensively studied in character animation [26, 27]. Our transition term adopts a lightweight boundary-consistency cost to penalize abrupt changes in the displacement of joints between consecutive segments of the proposed dance (Figure 3). Since our segments are derived from real motion data, we do not impose penalties for the transitions between poses within the segments. In other words, we only consider the cost of transitioning between every pair of adjacent segments (s_i, s_{i+1}).

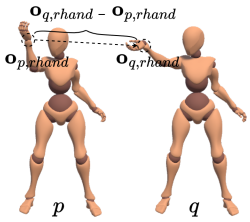


Figure 4: The transition cost computation between poses p and q for the hand joint.

Joint Displacement. We define the displacement between any pair of poses (p, q) as the average Euclidean distance between their corresponding joint positions:

$$\Delta(p, q) = \frac{1}{|\mathcal{J}|} \sum_{j \in \mathcal{J}} \|\mathbf{o}_{p,j} - \mathbf{o}_{q,j}\| \tag{3}$$

where $\mathbf{o}_{p,j}$ $\mathbf{o}_{q,j}$ denote the 3D position vectors of joint j in poses p and q, respectively. This metric captures the overall spatial deviation between two poses based on joint configurations. We illustrate how we compute the joint displacement in Figure 4.

Transition Cost. Finally, we compute the transition cost based on the average joint displacement between the final pose of one segment and the initial pose of the next. We do this by computing the average joint displacement between the last pose of the first segment s_i and the first pose of the next segment s_{i+1} for every pair of consecutive segments (s_i, s_{i+1}):

$$C_{\mathrm{T}}(\mathcal{S}) = \frac{1}{n-1}\left(1 - \exp\left(-\frac{\sum_{(p,q)} \Delta(p, q)^2}{2\sigma_{\mathrm{T}}^2}\right)\right) \tag{4}$$

where n refers to the number of dance segments in current generated sequence. For each consecutive segment pair (s_i, s_{i+1}) in sequence, we select the last pose from segment s_i as p and the first pose from segment s_{i+1} as q, $\Delta(p, q)$ is joint displacement between p and q as defined in Equation 3, and σ_{T} controls the sensitivity to displacement. We use linear interpolation for translations and log-space interpolation [16] for rotations within a short window around each segment boundary when animating entire dance sequence. Essentially, this cost term penalizes large boundary mismatches between consecutive segments, which would otherwise require substantial interpolation and lead to visually abrupt transitions.

5.4 Variety Cost

The variety cost encourages global segment-level diversity across the routine, promoting dance sequences with varying levels of memorability and cognitive complexity. It penalizes similarity between adjacent segments based on extracted motion features.

Feature Extraction. We first compute the feature vector for each pose in our segments. To capture energy expenditure in poses, we utilize the kinetic energy parameters from Onuma et al. [46]. Each pose contributes three energy parameters per joint, resulting in a feature vector of size 72 across all 24 joints. Additionally, we incorporate qualitative geometric features from Müller et al. [41] to capture spatio-temporal differences in dance segments. These manual features, including joint angles and distances, yield an additional 25-dimensional vector. By concatenating the kinetic and geometric features, each pose is represented by a 97-dimensional vector.

To compute the feature vector for a segment, we concatenate the features of all poses within the segment. Since each segment contains 120 poses, the resulting feature vector is $\mathbf{f}_s \in \mathbb{R}^{11640}$.

Segment Diversity. We can use the cosine similarity of their feature vectors to measure the difference between any pair of segments. Therefore, we define the variety of dance by computing the average divergence between all segments in proposed dance $\mathcal{S}$:

$$\varphi(\mathcal{S}) = 1 - \frac{1}{N} \sum_{(s,s')} cos(\mathbf{f}_s, \mathbf{f}_{s'}) \tag{5}$$

where $N = \binom{n}{2}$ represents the number of possible combinations of segment pairs that can be formed from a dance consisting of n segments. We compute the sum over all possible combination pairs of segments s and s' in $\mathcal{S}$. $\mathbf{f}_s$ and $\mathbf{f}_{s'}$ are the extracted feature vectors of the segments s and s'. A higher φ indicates an overall larger distance between the feature vectors and thus measures global segment-level diversity across all pairs of the dance.

Variety Cost. To encourage diverse choreography, we define a segment-level variety cost based on the average dissimilarity between segments. Segment diversity $\varphi(\mathcal{S})$ is computed as the mean cosine divergence between all segment feature pairs (Equation 5). To control this diversity, we compare it with a target value ρ^{V}:

$$C_{\mathrm{V}}(\mathcal{S}) = 1 - \exp\left(-\frac{(\varphi(\mathcal{S}) - \rho^{\mathrm{V}})^2}{2\sigma_{\mathrm{V}}^2}\right) \tag{6}$$

Here, σ_{V} controls the tolerance to deviations from the desired variety level.

This goal-based formulation enables us to control the diversity of movements within the dance routine. Unlike transition cost, which minimizes local inconsistency to ensure temporal smoothness, the variety cost controls the structure of the choreography.

5.5 Intensity Costs

The physical intensity of movements is approximated using the linear and rotational kinetic energy of each joint across poses [56]. We denote these as C_{LIN} and C_{ROT} for the linear and rotational intensity cost terms in Equation 1.

5.5.1 Linear Intensity

We compute this intensity cost by estimating the kinetic energy resulting from the joints' global translation [15, 56].

Linear Velocity. The linear velocity of joint j at any pose p is computed as the positional difference between consecutive poses:

$$\mathbf{v}_{p,j} = \frac{\mathbf{o}_{p,j} - \mathbf{o}_{q,j}}{\tau(p,q)} \tag{7}$$

where q is the pose preceding p in our dance $\mathcal{S}$, while $\mathbf{o}_{p,j}$ and $\mathbf{o}_{q,j}$ are the 3D positions of joint j in poses p and q respectively. $\tau(p,q)$ returns the time it takes to transition from the pose q to the pose p. $\tau(.,.)$ consistently returns 3.3×10^{-2} for all pose combinations, as all dances in our library are uniformly sampled at 30 FPS.

Translational Kinetic Energy. The translational kinetic energy measures the spatial movement of joints [56, 57]:

$$E_{\text{L}}(p) = \frac{1}{2|\mathcal{J}|} \sum_{j \in \mathcal{J}} m_j \left\| \mathbf{v}_{p,j} \right\|^2 \tag{8}$$

where $\mathbf{v}_{p,j}$ is the linear velocity of joint j at pose p (Equation 7) while m_j is the mass ratio assigned to the joint j. The joint mass ratios m_j were assigned based on anthropometric data [56]. $E_{\text{L}}(p)$ is commonly used in motion analysis research as a kinematics-based proxy for physical intensity [46]. The specific values we used are provided in supplementary material.

Linear Intensity Cost. The linear intensity cost aims to match the computed energy of the dance to a target value:

$$C_{\text{LIN}}(\mathcal{S}) = 1 - \exp\left(-\frac{(\sum_p E_{\text{L}}(p) - \rho^{\text{L}})^2}{2\sigma_{\text{L}}^2} \right) \tag{9}$$

where $p \in \mathcal{S}$ refers to a pose in the dance sequence $\mathcal{S}$, $E_{\text{L}}(p)$ is the linear kinetic energy of pose p, ρ^{L} is the target total linear energy, and σ_{L} controls the tolerance for deviations from the target ρ^{L}.

5.5.2 Rotational Intensity

We compute the rotational intensity using the local angular motion of joints [58, 65].

Angular Rotation. The angular rotation of joint j at any pose p is computed as the relative rotational difference between consecutive poses:

$$\omega_{p,j} = \frac{\log\left(\mathbf{R}_{q,j}^{-1} \mathbf{R}_{p,j}\right)}{\tau(p,q)} \tag{10}$$

where q is the pose preceding p in our generated dance $\mathcal{S}$, $\mathbf{R}_{p,j} \in SO(3)$ and $\mathbf{R}_{q,j} \in SO(3)$ denote the local rotation matrices of joint j (with respect to its parent) at poses p and q, respectively. $\text{Log}(\cdot)$ maps a relative rotation in $SO(3)$ to its corresponding rotation vector in the tangent space [16]. $\tau(p,q)$ returns the time it took to transition from pose q to pose p in seconds.

Rotational Kinetic Energy. The relative kinetic energy captures the internal angular motion of the joints [2, 65]:

$$E_{\text{R}}(p) = \frac{1}{2|\mathcal{J}|} \sum_{j \in \mathcal{J}} a_j m_j (k_j L_j)^2 \left\| \boldsymbol{\omega}_{p,j} \right\|^2 \tag{11}$$

where $\boldsymbol{\omega}_{p,j}$ is the angular velocity proxy of joint j at pose p (Equation 10). k_j is the radius of gyration coefficient and L_j is the segment length, derived from SMPL node distances scaled to match anthropometric data of average adult males [56]. Here, $a_j \in [0,1]$ is a joint-specific mask that scales the contribution of joint j to the energy. The specific values of m_j and k_j we used are provided in supplementary material.

Rotational Intensity Cost. Similarly, the relative intensity cost compares the rotational kinetic energy to a target ρ^{R}:

$$C_{\text{ROT}}(\mathcal{S}) = 1 - \exp\left(-\frac{(\sum_p E_{\text{R}}(p) - \rho^{\text{R}})^2}{2\sigma_{\text{R}}^2} \right) \tag{12}$$

where $p \in \mathcal{S}$ refers to a pose in the dance $\mathcal{S}$, $E_{\text{R}}(p)$ is the rotational kinetic energy of pose p, and σ_{R} controls the tolerance for deviations from the target ρ^{R}.

5.6 Optimization

Our optimizer synthesizes personalized dance routines by minimizing the total cost $C_{\text{Total}}(S)$ (Equation 1). Each dance S consists of multiple segments selected from the dance library (Section 4). We place no explicit limit on how often segments can be selected, relying instead on the variety cost to maintain diversity.

To efficiently explore different segment combinations, we use reversible-jump Markov chain Monte Carlo (RJ-MCMC) [18] guided by simulated annealing to refine the solutions. We define a Boltzmann-like objective function:

$$f(\mathcal{S}) = \exp\left(-\frac{C_{\text{Total}}(\mathcal{S})}{t} \right), \tag{13}$$

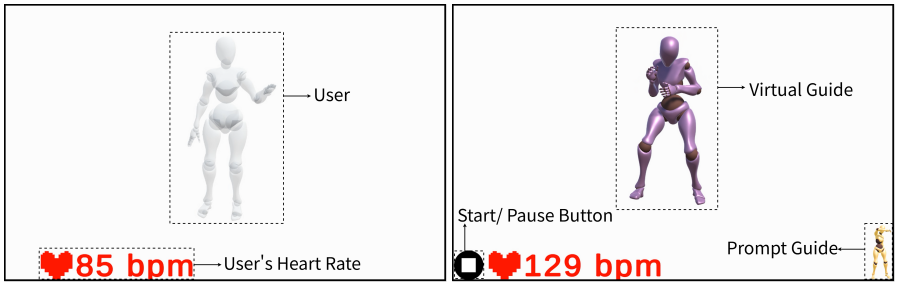
where t is the annealing temperature. Higher temperatures encourage broad exploration early in optimization. As t decreases, the optimizer increasingly favors improvements that reduce the total cost. Initially, we set the temperature $t = 1.0$. Then, the temperature decreases by 0.01 every 1000 iterations until it reaches zero, progressively shifting from exploration to refinement. Optimization terminates once the change in total cost falls below 3% over 50 iterations. Since our RJ-MCMC optimization is stochastic, the system can produce multiple distinct dance routines under the same target settings by varying the initialization.

Optimization Steps. At each iteration, we perform one of the three operations: remove a random dance segment within the routine, add a randomly sampled segment from the library to the end of the routine, or replace a randomly selected segment in the solution with a random segment from the library. The probability of performing the remove operation was set as 20%. The probability of the other two was set to 40%, favoring them over removal.

Parameter Settings. We set the weights for the cost terms as $\lambda_{\text{D}} = \lambda_{\text{LIN}} = \lambda_{\text{ROT}} = 1.0$ in Equation 1 for all our results. For the transition term, we set $\lambda_{\text{T}} = 1.0$ to enable the transition cost or $\lambda_{\text{T}} = 0$ to disable it, allowing us to directly compare routines with and without transition constraints in our experiments.

The default target duration ρ^{D} is set to 32 seconds, corresponding to eight sets of eight-beat movements, a standard measure in choreography [43, 62]. To obtain the parameters for the other costs, we randomly sampled 1,000 segments of 32-second duration from our preprocessed data before segmentation.

We invited 3 experienced dancers to classify the physical intensity (*Light* or *Intense*) and difficulty (*Easy* or *Hard*) of the 1,000 sampled segments. We used their feedback to compute the target values used in our user study (Section 6) and to generate the dance routines in Figures 6 & 3. For each annotated category, we calculated the mean values of translational kinetic energy, rotational kinetic energy, and motion variety across all samples, and used these averages as the target values. For the *Light* intensity conditions, the targets were set to $\rho^{\text{L}} = 56.5$ and $\rho^{\text{R}} = 31.8$, while in the *Intense* conditions, they were $\rho^{\text{L}} = 147.2$ and $\rho^{\text{R}} = 106.5$. For variety, we used $\rho^{\text{V}} = 67.4$ for the *Easy* conditions and $\rho^{\text{V}} = 143.1$ for the



(a) Training (b) Performance

Figure 5: To evaluate and demonstrate the effectiveness of our framework, we developed a Unity-based dance exergame. The game consists of two main phases: (a) *Training Phase* and (b) *Performance Phase*. In the Training Phase, the player can observe their real-time motion and heart rate feedback while freely exploring and rehearsing a given dance. Once familiarized, they proceed to the Performance Phase where a central *Virtual Guide* performs the target choreography in real-time. A *Prompt Guide* displays upcoming poses at double speed relative to the Virtual Guide. In the upper-right corner, the game provides real-time visual feedback by triggering animations such as "Perfect" or "Good" based on the DTW distance between the player's current pose and the corresponding reference pose segment.

Hard conditions. For the dances shown in (Figure 3), we generated sequences under the *Light-Easy* setting and compared versions with and without the transition term by modifying λ_{T}. Figure 6 shows the generated sequences with transition enabled $\lambda_{\mathrm{T}} = 1$.

From the same sampled data, we computed the mean and standard deviations of motion diversity (Equation 5), translational kinetic energy (Equation 8), rotational kinetic energy (Equation 11), total joint displacement (Equation 3), and duration across the population for each segment to set the scale parameters of our cost functions. This allowed us to set the values of $\sigma_V = 48.3$, $\sigma_L = 88.4$, $\sigma_R = 48.0$, $\sigma_T = 0.03$. Since all sampled segments were 32 seconds long, the variance of duration is negligible; therefore, we set σ_D to a constant value of 1.0 for normalization purposes.

To verify our parameter settings for the average user's dance experience, we conducted a pilot study involving five participants. In this study, we tested three duration settings (16s, 32s, and 48s) while varying our aforementioned intensity and difficulty target values. Participants provided qualitative feedback on how well each configuration matched the described physical and cognitive intensity classifications. Results indicated that 16-second dances were perceived as too short to produce meaningful physical exertion, whereas 48-second sequences often led to fatigue and an increased number of movement errors. These observations confirmed our selection of 32 seconds for the dance durations shown to participants in our user study.

Animation. After the optimization, we apply log-space and linear interpolation to joint rotations and root translations between each of the dance segments in the generated sequence. Specifically, we interpolate only within a short window (10 frames) around each segment boundary to reduce visual discontinuities caused by endpoint mismatches. This ensures smoother transitions across the dance. The interpolated motion is then reconstructed into a playable dance animation using an inverse kinematics approach, which incorporates keypoint reprojection and smoothness regularization costs following Li et al. [35]. To render the animation, the reconstructed 3D positions and rotations are converted into the SMPL model format, which is widely supported by animation and modeling software such as Blender. Finally, our exergame presents the optimized dance as a full-body animation of a digital avatar (Figure 1).

Implementation We conducted the experiments using an Alienware Aurora R16 equipped with an Intel Core i7-13700F, 64GB of memory, and an RTX 4060 GPU. The optimization, data capture, and exergames (Figure 5) were implemented using Python 3.8. We captured participants' motion using the Femto Bolt camera and Orbbec SDK, and heart rate data using the Polar H10 Heart Rate Sensor. Our system performs optimization each time user updates dance specifications. For the eight conditions tested in our study, the average optimization time was approximately 51 seconds.

6 User Study

We conducted an ablation study to investigate the proposed costs and their influence on real-world physical and cognitive training. We also assessed the enjoyment level of physical activity using our exergame and compared it with other systems.

Participants. We recruited 20 participants (10 male, 10 female) from the university's student and staff population to participate in a study approved by the Institutional Review Board (IRB) of The University of Texas at Dallas (IRB-24-671). All participants provided informed consent. Their ages ranged from 21 to 40 years ($M = 25.3$, $SD = 6.3$), and their average weight was 73.5 kg ($SD = 8.9$). Among the participants, one had more than three years of dance experience, two had less than one year of experience, and the remaining seventeen had no prior dance experience.

Procedure. The study took place in a spacious indoor environment. Participants were fitted with a Polar H10 heart rate sensor and were asked to perform the dance routines. First, participants completed a training session to familiarize themselves with the system and gameplay mechanics (Figure 5a).

After the training session, participants completed eight conditions, consisting of first watching a dance to help with memorization, then performing it. In other words, we followed a $2 \times 2 \times 2$ within-subject study design where we varied our three optimization costs, *Intensity (Light, Intense)* $\times$ *Variety (Easy, Difficult)* $\times$ *Transition (Enabled, Disabled)*, to generate eight condition dances. The ordering of the dance conditions was set using a Latin Square to mitigate order effects. Participants were given a 5-minute rest period after each condition to prevent fatigue and reset their heart rate, and they were asked to complete the survey provided in the supplementary material. Figure 6 shows the dances generated while enabling the transition cost. Figure 3 shows the dances generated for varying the *Transition* condition under the *Light-Easy* condition.

After completing all eight dance conditions, participants completed a survey to evaluate their enjoyment of the system and provide their demographic information.

6.1 Results

Physical. We evaluate the effectiveness of our optimization at generating dances for various exercise goals.

The intensity costs strongly correlate with participants' heart rate. To show this, we conducted a Pearson's product-moment correlation test, which showed significant positive correlations between heart rate and the sum of the kinetic translational and rotational energies across conditions ($r = 0.883$, $p < 0.001$). Figure 6 shows participants' average heart rate under each condition.

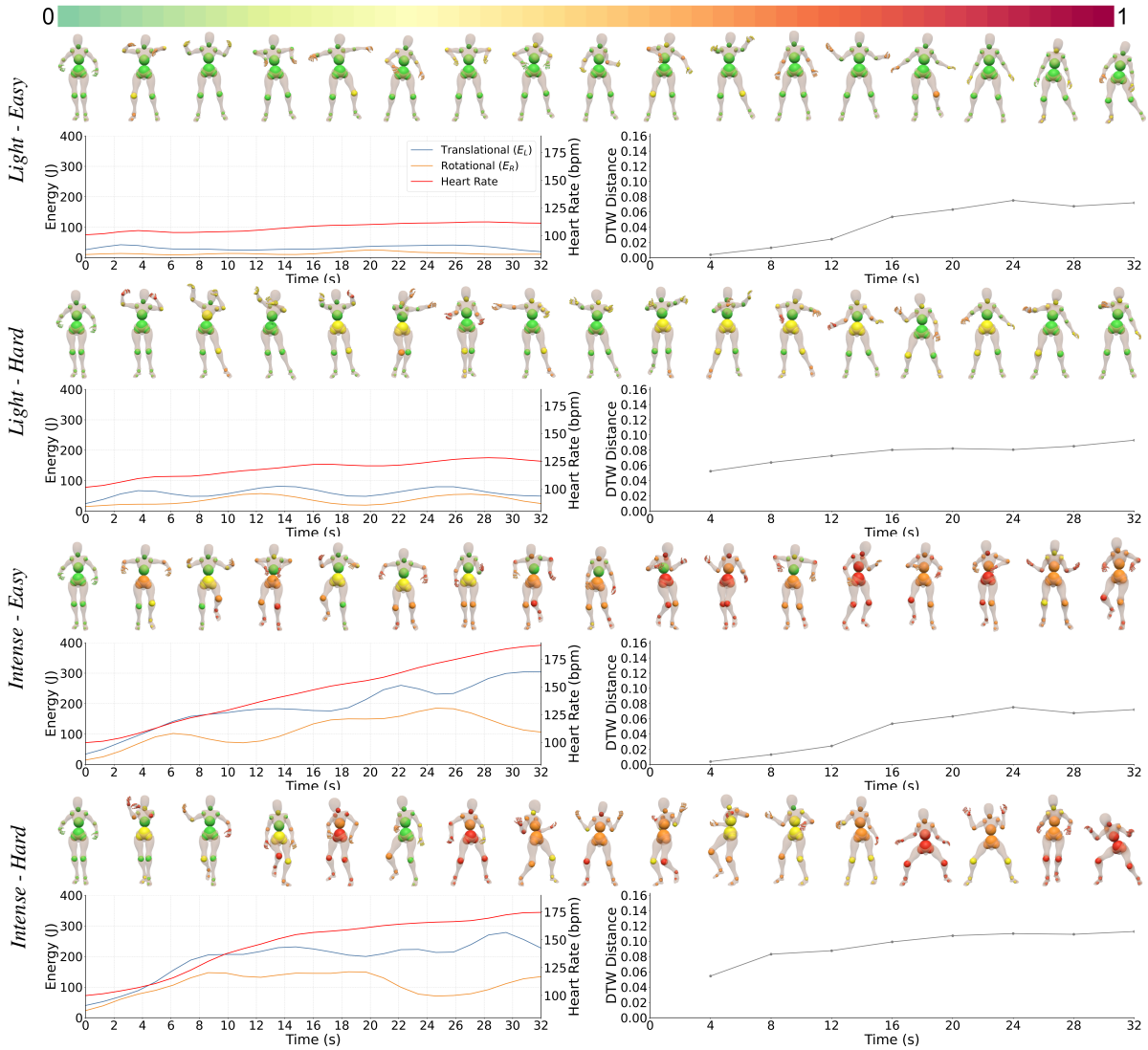


Figure 6: Visualization of different training goal dances with joint-level intensities normalized across joints and mapped to a color scale.

We also found that the proportion of time spent at vigorous intensity exercise zones was higher under the *Intense* conditions. According to the American Heart Association (AHA), vigorous-intensity exercise corresponds to 70% or higher of an individual's estimated maximum heart rate ($208 - 0.7 \times$ age) [52]. We analyzed each participant's heart rate across four dance categories: *Light-Easy*, *Light-Hard*, *Intense-Easy*, and *Intense-Hard*. Averaged across participants, the proportions of time spent at or above the vigorous-intensity were: 0.00% for *Light-Easy*, 2.56% for *Light-Hard*, 66.17% for *Intense-Easy*, and 73.67% for *Intense-Hard*.

In addition, a Friedman test revealed a significant effect of dance condition on average heart rate ($\chi^2(3) = 30.00, p < 0.001, W = 0.50$). Post-hoc Wilcoxon signed-rank tests showed that all condition pairs differed significantly ($p < 0.01$), indicating that our intensity costs influenced the physical training intensity.

These results confirm that the physical intensity costs of our dance routines reliably reflect participants' physiological responses. Therefore, controlling our proposed cost can effectively adjust the physical intensity of generated dances.

Cognitive. We also evaluated whether adjusting our variety cost's goal ρ^{V} has any effect on the memorability of the dance routines.

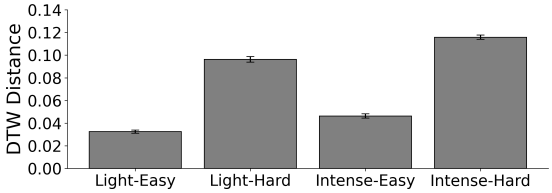


Figure 7: Mean DTW distance across participants for four conditions.

Performance measurements like accuracy and error rate have been shown to be good estimates of memory performance [49]. Therefore, we computed the Dynamic Time Warping (DTW) [11] distance between the generated dance shown to participants and their recorded motion to measure their performance and assess their memorization (Figure 6).

We found a strong correlation between our variety cost and the DTW distance between participants' performance and the generated reference dance across all conditions ($r = 0.897, p = 0.031$). This suggests that manipulating our variety cost affects participants' performance accuracy. Figure 6 shows participants' average DTW distance over time, while Figure 7 shows the average distance computed across participants.

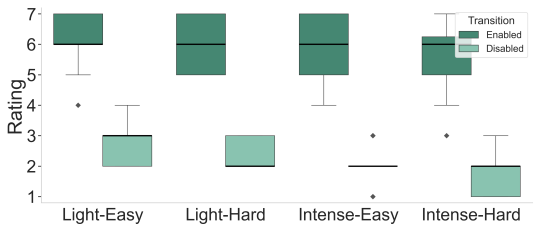


Figure 8: Participants' ratings of the dance smoothness.

We also found a significant difference in the participants' performance accuracy between the *Easy* and *Hard* conditions. A Friedman test revealed a significant difference in the DTW distance between conditions ($\chi^2(3) = 30.0, p < 0.001, W = 0.50$). Furthermore, a post-hoc Wilcoxon signed-rank test showed significant differences between all condition pairs ($p < 0.01$), including *Light-Easy vs. Intense-Easy*, *Light-Easy vs. Light-Hard*, *Light-Easy vs. Intense-Hard*, *Intense-Easy vs. Light-Hard*, *Intense-Easy vs. Intense-Hard*, and *Light-Hard vs. Intense-Hard*, confirming that our variety cost affects participants' performance accuracy.

Our results show that our proposed variety cost can control the variation of movements within the dance, which also indirectly affects the difficulty of performing them. In other words, this cost allows us to control how challenging the dances are to memorize for cognitive training.

Transition. We examined whether the transition cost influenced the perceived smoothness of generated dances based on user ratings (Figure 8). Participants evaluated eight conditions defined by different intensity settings and the presence or absence of the transition cost. Participants rated the conditions from 1 (strongly disagree) to 7 (strongly agree), in response to questions regarding dance smoothness. A Friedman test revealed significant differences across conditions ($\chi^2(7) = 18.53, p < 0.01$). Post-hoc Wilcoxon signed-rank tests with Bonferroni correction showed that, across all conditions, dances generated with the transition cost enabled were consistently rated as smoother than those with the cost disabled ($p < 0.01$). These findings demonstrate that the transition cost serves as an effective control parameter for regulating the smoothness of generated dances. In the previous analyses of physical and cognitive intensity, only conditions with the transition cost enabled were included, as disabling it disrupted motion continuity, prevented participants from completing the movements, and consequently confounded both heart rate and DTW distance measurements.

Physical Activity Enjoyment Rating. To evaluate perceived enjoyment, we administered the Physical Activity Enjoyment Scale (PACES), an 18-item questionnaire using a 7-point Likert scale designed to assess enjoyment of physical activity [24]. Participants gave an average PACES percentage score of 85% to our system. Average and standard deviation of ratings are provided in the supplementary material.This score indicates a high level of enjoyment and compares favorably to other widely used commercial exergames. For example, Graves et al. [17] reported PACES scores of 67% for Wii Yoga, 74% for Wii Muscle, and 85% for Wii Aerobic.

In comparison to the system proposed by Zhang et al. [66], which reported a PACES percentage score of 79%, our system demonstrates improved enjoyment. We attribute this to two key factors. First, the dance-based movements in our system are more natural and closely resemble real-world choreography compared to the discrete poses in Zhang et al.'s system. Second, the increased complexity of dance movements presents a moderate physical and cognitive challenge, which contributes to heightened engagement.

7 Experiments

Our optimization-based approach allows us to control the intensity levels for certain joints, which is particularly useful in the case of injury or when there's a desire to focus exercise intensity on specific areas. Additionally, this approach enables us to control the physical intensity levels and variety of movements within the dance routine. The parameter settings used to generate the dances are provided in the supplementary material.

7.1 Constraining Joint Intensities

Our system allows for some joint-level intensity control by introducing a joint-specific mask, enabling the generation of dance routines that reduce exertion in targeted body regions while maintaining high intensity elsewhere. This is valuable in rehabilitation or physical therapy scenarios where injuries require the therapist to provide selective muscle training [22]. While the base objective (Equation 11) computes the rotational kinetic energy of the full body, a joint mask $a_j \in [0, 1]$ is introduced to scale the contribution of each joint j. Setting $a_j = 0$ fully constrains joint j, while $a_j = 1$ leaves it unconstrained. Setting the mask to a value between zero and one enables us to control the joint's influence on the overall intensity.

To illustrate this, we generated a dance routine with constrained motion at the right upper limb while preserving high intensity and high variety in the rest of the body (Figure 9). Specifically, We set $a_{\text{right elbow}} = 0$ and $a_{\text{right shoulder}} = 0$ for the right elbow and right shoulder. As a result, the optimizer avoided dance segments with large rotational motion in the constrained joints, while favoring energetic movements in the remaining joints. Most movements in the routine involve minimal rotation around the right elbow, demonstrating our framework's ability to generate dances that respect joint intensity constraints. We plotted the sequence's translational and rotational kinetic energies, along with the heart rate captured from the user over time.

7.2 Varying Physical and Cognitive Intensity

While recent learning-based methods are effective at generating aesthetically pleasing motions, they are unable to offer control over the changes in physical intensity or variety of movements within the dance. This is unlike our optimization-based framework, which allows us to vary the training goals within the dance, enabling the creation of dances that vary in physical and cognitive demands across different segments.

We can achieve this by adjusting our costs to minimize the values relative to a list of user-specified goals. For example, we can modify our previous linear intensity cost (Equation 9) as follows:

$$C_{\text{SLIN}}(\mathcal{S}) = 1 - \exp\left(-\frac{\sum_s (E_{\text{L}}(s) - \rho_s^{\text{L}})^2}{2\sigma_{\text{L}}^2}\right) \qquad (14)$$

where $E_{\text{L}}(s)$ computes the translational kinetic energy for each segment s and ρ_s^{L} is the translational energy goal for segment s.

We demonstrate this capability by generating two 128-second dance sequences by controlling the intensity and variety targets (Figure 10). In Figure 10a, we generate the dance sequence by alternating the kinetic energy between 32 second bursts of intense activity and rest. This design achieves a high-intensity interval training (HIIT) routine, which has been shown to improve training experience more effectively when compared to continuous high-intensity exercise [14, 55]. We can observe an interval training effect on the user's captured heart rate during the dance, where intense periods elevate the heart rate and lighter periods allow for partial recovery, lowering the heart rate. Similarly, generate the dance sequence shown in Figure 10b by increasing and decreasing the variety targets every 32 seconds. During the performance of the dance, the DTW distance between the user's captured motion and the generated dance fluctuates between low and high values. This alternation indicates changes in the difficulty of accurately performing the dance as we move between simple/repetitive and complex subsequences.

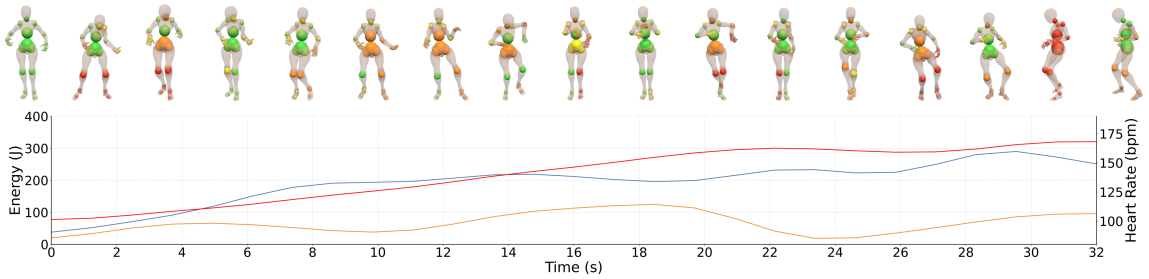


Figure 9: Kinetic-energy profiles and participant heart rate over time for a routine generated with right-arm joint-intensity constraints.

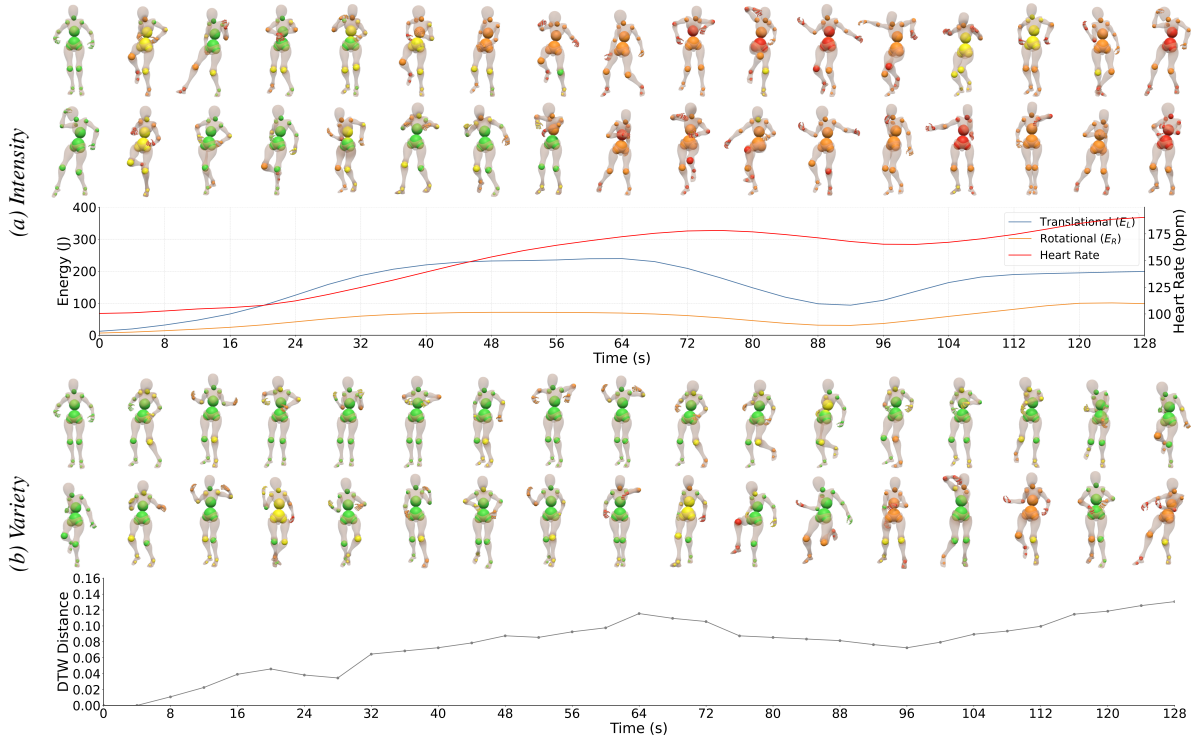


Figure 10: Varying the dance sequence's (a) intensity and (b) variety targets.

8 Limitations and Future Work

Our intensity terms are designed as controllable motion-based proxies that support efficient optimization and user-facing control, but they simplify several aspects of real-world training. The physical intensity objectives are kinematics-derived (from joint translations and relative joint rotations on $SO(3)$), yet they do not explicitly model stance/swing phases, ground-reaction forces, or joint torques; thus, they should be interpreted as motion-level effort proxies rather than biomechanically complete measures. In addition, the segment-based stochastic optimization can yield different routines under the same targets through different initializations, but exhaustive coverage of solutions is not guaranteed under fixed hyperparameters and finite compute budgets. Finally, the current optimization does not enforce beat-level music synchronization, so musical alignment may be inconsistent when segments are sampled independently.

Cross-segment Joint Blending. Currently, our system selects "full-body" motion segments from the dance library. One promising direction is to splice joint-specific motion from one segment into another, enabling more flexible choreography generation. For example, motion from the right arm can be preserved while freezing the left arm by directly substituting joint trajectories. Achieving this requires modeling inter-joint constraints and introducing joint-level consistency costs to maintain biomechanical plausibility during segment recombination.

Incorporating Music. Recent work has explored music-conditioned and rhythm-aware dance generation [3, 13, 60]. Our current framework does not optimize dance routines for musical structure. As segments are sampled independently, we may lose temporal consistency with the music's rhythm.. In future work, we aim to introduce music-aware cost terms that enforce beat synchronization during optimization, such as Mel-Frequency Cepstral Coefficients, beat histograms, onset strength, and tempograms [10,40]. These features can guide the optimizer to synchronize movements with beats, align transitions with tempo changes, and enhance the emotional expression, aesthetic quality, and artistic coherence of the generated dances.

Real-Time Adaption to Physiological Feedback. Our current design generates dance routines with predefined intensity settings. However, real-time physiological feedback, such as heart rate data, offers the opportunity to adapt the dance to keep users within their target heart rate zones. In future work, we plan to integrate real-time heart rate monitoring within the optimization loop to enable real-time dance generation based on current exertion levels. This could support more effective cardiovascular training, personalized pacing, and safe yet challenging dance sessions.

REFERENCES

[1] D. S. Alexiadis, P. Daras, P. Kelly, N. E. O'Connor, and T. Boubekeur. Evaluating a dancer's performance using kinect-based skeleton tracking. In *Proceedings of the 19th ACM International Conference on Multimedia*, pp. 659–662, 2011. doi: 10.1145/2072298.2072412 1

[2] O. Arikan, D. A. Forsyth, and J. F. O'Brien. Motion synthesis from annotations. In *ACM SIGGRAPH 2003 Papers*, pp. 402–408. 2003. 5

[3] H. Y. Au, J. Chen, J. Jiang, and Y. Guo. Choreograph: Music-conditioned automatic dance choreography over a style and tempo consistent dynamic graph. In *Proceedings of the 30th ACM International Conference on Multimedia*, pp. 3917–3925, 2022. 9

[4] C. Basak, W. R. Boot, M. W. Voss, and A. F. Kramer. Can training in a real-time strategy video game attenuate cognitive decline in older adults? *Psychology and Aging*, 23(4):765–777, 2008. Cited by: 928. doi: 10.1037/a0013494 1

[5] C. Basak, S. E. Qin, and M. A. O'Connell. Differential effects of cognitive training modules on healthy aging and mild cognitive impairment: A comprehensive meta-analysis of randomized controlled trials. *Psychology and Aging*, 35(2):220–249, 2020. 1

[6] C. G. Bennett, R. P. Guttmann, M. E. Hackney, R. Amin, and S. Weaver. Impacts of adapted dance on mood and physical function among persons living with alzheimer's disease. *Journal of Alzheimer's Disease*, 105(4):1148–1159, 2025. 2

[7] P. Caserman, S. Liu, and S. Göbel. Full-body motion recognition in immersive-virtual-reality-based exergame. *IEEE Transactions on Games*, 14(2):243–252, 2021. 2

[8] K. Chen, Z. Tan, J. Lei, S.-H. Zhang, Y.-C. Guo, W. Zhang, and S.-M. Hu. Choreomaster: Choreography-oriented music-driven dance synthesis. *ACM Transactions on Graphics (TOG)*, 40(4):1–13, 2021. doi: 10.1145/3450626.3459932 1, 2

[9] A. Correia and L. A. Alexandre. Music to dance as language translation using sequence models. *arXiv preprint arXiv:2403.15569*, 2024. 2

[10] D. P. Ellis. Beat tracking by dynamic programming. *Journal of New Music Research*, 36(1):51–60, 2007. doi: 10.1080/09298210701653344 9

[11] S. Ferguson, E. Schubert, and C. J. Stevens. Dynamic dance warping: Using dynamic time warping to compare dance movement performed under different conditions. In *Proceedings of the 2014 International Workshop on Movement and Computing (MOCO '14)*, MOCO '14, pp. 94–99. Association for Computing Machinery, Paris, France, June 2014. doi: 10.1145/2617995.2618012 7

[12] Z. Gao, J. E. Lee, and Z. Pope. Exergaming, active video games, and play: An interdisciplinary overview. *Games for Health Journal*, 7(1):1–12, 2018. 1

[13] A. Ghosh, B. Zhou, R. Dabral, J. Wang, V. Golyanik, C. Theobalt, P. Slusallek, and C. Guo. Duetgen: Music driven two-person dance generation via hierarchical masked modeling. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*, pp. 1–11, 2025. 9

[14] M. J. Gibala and S. L. McGee. Metabolic adaptations to short-term high-intensity interval training: a little pain for a lot of gain? *Exercise and sport sciences reviews*, 36(2):58–63, 2008. 8

[15] H. Goldstein, C. Poole, J. Safko, and S. R. Addison. *Classical mechanics*. American Association of Physics Teachers, 2002. 5

[16] F. S. Grassia. Practical parameterization of rotations using the exponential map. *Journal of graphics tools*, 3(3):29–48, 1998. 4, 5

[17] L. E. Graves, N. D. Ridgers, K. Williams, G. Stratton, G. Atkinson, and N. T. Cable. The physiological cost and enjoyment of wii fit in adolescents, young adults, and older adults. *Journal of Physical Activity and Health*, 7(3):393–401, 2010. 8

[18] P. J. Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732, 1995. doi: 10.1093/biomet/82.4.711 5

[19] H. Han, K. Jung, and S. H. Yoon. Choreocraft: Insitu crafting of choreography in virtual reality through creativity support tool. In *CHI Conference on Human Factors in Computing Systems (CHI '25)*, p. 21. ACM, New York, NY, USA, April 2025. doi: 10.1145/3706598.3714220 1, 2

[20] Harmonix Music Systems. Dance Central. https://www.harmonixmusic.com/games/dance-central, 2010. Accessed: 2024-05-05. 1

[21] P. W.-N. Hwang and K. L. Braun. The effectiveness of dance interventions to improve older adults' health: A systematic literature review. *Alternative Therapies in Health and Medicine*, 21(5):64–70, 2015. 1

[22] D. U. Jette, N. K. Latham, R. J. Smout, J. Gassaway, M. D. Slavin, and S. D. Horn. Physical therapy interventions for patients with stroke in inpatient rehabilitation facilities. *Physical therapy*, 85(3):238–248, 2005. 8

[23] K. Kakoutopoulos, E. Drakakis, A. Papadopoulou, and C. Goumopoulos. Feasibility of augmented reality-based cognitive training for older adults: The marketmind ar approach. *Sensors*, 25(7):2081, 2025. doi: 10.3390/s25072081 2

[24] D. Kendzierski and K. J. DeCarlo. Physical activity enjoyment scale: Two validation studies. *Journal of Sport and Exercise Psychology*, 13(1):50–64, 1991. 8

[25] M. Kim, J. Hwang, S.-H. Lee, J. Kim, and J. Noh. Tiling motion patches. In *Proceedings of the Eurographics/ACM SIGGRAPH Symposium on Computer Animation*, pp. 117–126. Eurographics Association, 2012. 2

[26] L. Kovar, M. Gleicher, and F. Pighin. Motion graphs. *ACM Trans. Graph.*, 21(3):473–482, July 2002. doi: 10.1145/566654.566605 4

[27] L. Kovar, M. Gleicher, and F. Pighin. *Motion Graphs*. Association for Computing Machinery, New York, NY, USA, 1 ed., 2023. 4

[28] N. Le, T. Do, K. Do, H. Nguyen, E. Tjiputra, Q. D. Tran, and A. Nguyen. Controllable group choreography using contrastive diffusion. *ACM Transactions on Graphics (TOG)*, 42(6):1–14, 2023. 2

[29] N. Le, T. Pham, T. Do, E. Tjiputra, Q. D. Tran, and A. Nguyen. Music-driven group choreography. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12345–12354, 2023. 2

[30] R. Li, S. Yang, D. A. Ross, and A. Kanazawa. Ai choreographer: Music conditioned 3d dance generation with aist++. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 13401–13412, 2021. 2

[31] R. Li, H. Zhang, Y. Zhang, Y. Zhang, J. Guo, Y. Zhang, X. Li, and Y. Liu. Lodge++: High-quality and long dance generation with vivid choreography patterns. *arXiv preprint arXiv:2410.20389*, 2024. 2

[32] R. Li, Y. Zhang, Y. Zhang, H. Zhang, J. Guo, Y. Zhang, Y. Liu, and X. Li. Lodge: A coarse to fine diffusion network for long dance generation guided by the characteristic dance primitives. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1524–1534, 2024. 2

[33] R. Li, J. Zhao, Y. Zhang, M. Su, Z. Ren, H. Zhang, Y. Tang, and X. Li. Finedance: A fine-grained choreography dataset for 3d full body dance generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10234–10243, 2023. 2

[34] W. Li, B. Xie, Y. Zhang, W. Meiss, H. Huang, and L.-F. Yu. Exertion-aware path generation. *ACM Trans. Graph.*, 39(4), Aug. 2020. doi: 10.1145/3386569.3392393 2

[35] Y.-L. Li, X. Wu, X. Liu, Z. Wang, Y. Dou, Y. Ji, J. Zhang, Y. Li, X. Lu, J. Tan, et al. From isolated islands to pangea: Unifying semantic space for human action understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16582–16592, 2024. 6

[36] W. Liu, Y. Zhang, B. Zhang, Q. Xiong, H. Zhao, S. Li, J. Liu, and Y. Bian. Self-guided dmt: exploring a novel paradigm of dance movement therapy in mixed reality for children with asd. *IEEE Transactions on Visualization and Computer Graphics*, 30(5):2119–2128, 2024. 2

[37] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black. SMPL: A skinned multi-person linear model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 34(6):248:1–248:16, Oct. 2015. 3

[38] Majesco Entertainment. Zumba Fitness. https://zumbafitnessgame.com, 2010. Accessed: 2024-05-05. 1

[39] F. F. Mueller, D. Edge, F. Vetere, M. R. Gibbs, S. Agamanolis, and B. Bongers. Designing sports: a framework for exertion games. Association for Computing Machinery, New York, NY, USA, 2011. doi:

10.1145/1978942.1979330 1, 2

[40] M. Müller. *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications*. Springer, 2015. 9

[41] M. Müller, T. Röder, and M. Clausen. Efficient content-based retrieval of motion capture data. *ACM Transactions on Graphics*, 24(3):677–685, 2005. doi: 10.1145/1073204.1073247 4

[42] V.-M. Nurkkala, J. Kalermo, and T. Jarvilehto. Development of exergaming simulator for gym training, exercise testing and rehabilitation. *Journal of Communication and Computer*, 11:403–411, 2014. 2

[43] F. Ofli, E. Erzin, Y. Yemez, and A. M. Tekalp. Learn2dance: Learning statistical music-to-dance mappings for choreography synthesis. *IEEE Transactions on Multimedia*, 14(3):747–759, 2012. doi: 10.1109/TMM.2011.2181492 3, 5

[44] E. F. Ogawa, T. You, and S. G. Leveille. Potential benefits of exergaming for cognition and dual-task function in older adults: A systematic review. *Journal of Aging and Physical Activity*, 24(2):332–336, 2016. doi: 10.1123/japa.2014-0267 2

[45] J. D. K. Ongchoco, M. M. Chun, and W. A. Bainbridge. What moves us? the intrinsic memorability of dance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 49(6):889–899, 2023. doi: 10.1037/xlm0001168 1

[46] K. Onuma, C. Faloutsos, and J. K. Hodgins. FMDistance: A fast and effective distance function for motion capture data. In *Proceedings of the Eurographics 2008 - Short Papers*, pp. 83–86. Eurographics Association, 2008. 4, 5

[47] B. Sarupuri, R. Kulpa, A. Aristidou, and F. Multon. Dancing in virtual reality as an inclusive platform for social and physical fitness activities: a survey. *The Visual Computer*, 40(12):4055–4070, 2024. 2

[48] G. Singh and W. Jakob. Mcmc: Bridging rendering, optimization and generative ai. In *SIGGRAPH Asia 2024 Courses*, SA Courses '24. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3680532.3689592 2

[49] E. T. Smith, P. Skolasinska, S. Qin, A. Sun, and C. Basak. Cognitive and structural predictors of novel task learning, and contextual predictors of time series of daily task performance during the learning period. *Frontiers in Aging Neuroscience*, 14:936528, 2022. doi: 10.3389/fnagi.2022.936528 7

[50] M. Soleimani, M. Ghazisaeedi, and S. Heydari. The efficacy of virtual reality for upper limb rehabilitation in stroke patients: a systematic review and meta-analysis. *BMC Medical Informatics and Decision Making*, 24(1):135, 2024. doi: 10.1186/s12911-024-02534-y 1, 2

[51] A. E. Staiano and S. L. Calvert. Exergames for physical education courses: Physical, social, and cognitive benefits. *Child development perspectives*, 5(2):93–98, 2011. 2

[52] H. Tanaka, K. D. Monahan, and D. R. Seals. Age-predicted maximal heart rate revisited. *Journal of the American College of Cardiology*, 37(1):153–156, 2001. doi: 10.1016/S0735-1097(00)01054-8 7

[53] Ubisoft. Just Dance 2023 Edition. https://www.ubisoft.com/en-us/game/just-dance/2023, 2022. Accessed: 2024-05-05. 1

[54] T. Wang, L. Li, K. Lin, Y. Zhai, C.-C. Lin, Z. Yang, H. Zhang, Z. Liu, and L. Wang. Disco: Disentangled control for realistic human dance generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9326–9336, 2024. 2

[55] K. S. Weston, U. Wisløff, and J. S. Coombes. High-intensity interval training in patients with lifestyle-induced cardiometabolic disease: a systematic review and meta-analysis. *British journal of sports medicine*, 48(16):1227–1234, 2014. 8

[56] D. A. Winter. *Biomechanics and Motor Control of Human Movement*. Wiley, Hoboken, NJ, USA, 4th ed., 2009. doi: 10.1002/9780470549148 5

[57] A. Witkin and M. Kass. Spacetime constraints. In *Proceedings of the 15th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '88)*, pp. 159–168, 1988. doi: 10.1145/54852.378507 5

[58] K. Yamane and J. K. Hodgins. Simultaneous tracking and balancing of humanoid robots for imitating human motion capture data. *IEEE Transactions on Robotics*, 25(3):502–517, 2009. 5

[59] K. Yang, X. Tang, R. Diao, H. Liu, J. He, and Z. Fan. Codancers: Music-driven coherent group dance generation with choreographic unit. In *Proceedings of the 2024 International Conference on Multimedia Retrieval*, ICMR '24, p. 675–683. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3652583.3657998 2

[60] K. Yang, X. Tang, Z. Peng, Y. Hu, J. He, and H. Liu. Megadance: Mixture-of-experts architecture for genre-aware 3d dance generation. *arXiv preprint arXiv:2505.17543*, 2025. 9

[61] K. Yang, X. Tang, H. Wu, B. Qin, H. Liu, J. He, and Z. Fan. Cohedancers: Enhancing interactive group dance generation through music-driven coherence decomposition. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 6663–6671, 2025. 2

[62] K. Yang, X. Zhou, X. Tang, R. Diao, H. Liu, J. He, and Z. Fan. Beatdance: A beat-based model-agnostic contrastive learning framework for music-dance retrieval. In *Proceedings of the 2024 International Conference on Multimedia Retrieval*, ICMR '24, p. 11–19. Association for Computing Machinery, New York, NY, USA, 2024. doi: 10.1145/3652583.3658045 3, 5

[63] L. Yang, H. Wang, W. Yang, and J. Sun. Synthesizing open worlds with constraints using locally annealed reversible jump mcmc. *ACM Transactions on Graphics (TOG)*, 31(4):1–10, 2012. 2

[64] Z. Yang, Y.-H. Wen, S.-Y. Chen, X. Liu, Y. Gao, Y.-J. Liu, L. Gao, and H. Fu. Keyframe control of music-driven 3d dance generation. *IEEE Transactions on Visualization and Computer Graphics*, 30(7):3474–3486, 2023. 2

[65] V. M. Zatsiorsky. *Kinetics of Human Motion*. Human Kinetics, Champaign, IL, USA, 2002. 5

[66] Y. Zhang, B. Xie, H. Huang, E. Ogawa, T. You, and L.-F. Yu. Pose-guided level design. In *Proc. CHI Conf. Hum. Factors Comput. Syst.*, pp. 1–13, 2019. doi: 10.1145/3290605.3300784 1, 2, 8

[67] S. Zhu, Y. Sui, Y. Shen, Y. Zhu, N. Ali, C. Guo, and T. Wang. Effects of virtual reality intervention on cognition and motor function in older adults with mild cognitive impairment or dementia: A systematic review and meta-analysis. *Frontiers in Aging Neuroscience*, 13:586999, 2021. doi: 10.3389/fnagi.2021.586999 2