Adult2Child: Dynamic Scaling Laws to Create Child-Like Motion

Yuzhu Dong
Department of CISE
University of Florida
Gainesville, Florida

Sachin Paryani
Department of CISE
University of Florida
Gainesville, Florida

Neha Rana
Department of CISE
University of Florida
Gainesville, Florida

Aishat Aloba
Department of CISE
University of Florida
Gainesville, Florida

Lisa Anthony
Department of CISE
University of Florida
Gainesville, Florida

Eakta Jain
Department of CISE
University of Florida
Gainesville, Florida

ABSTRACT

Child characters are widely used in animations and games; however, child motion capture databases are less easily available than those involving adult actors. Previous studies have shown that there is a perceivable difference in adult and child motion based on point light displays, so it may not be appropriate to just use adult motion data on child characters. Due to the costs associated with motion capture of child actors, it would be beneficial if we could create a child motion corpus by translating adult motion into child-like motion. Previous works have proposed dynamic scaling laws to transfer motion from one character to its scaled version. In this paper, we conduct a perception study to understand if this procedure can be applied to translate adult motion into child-like motion. Viewers were shown three types of point light display videos: adult motion, child motion, and dynamically scaled adult motion and asked to identify if the translated motion belongs to a child or an adult. We found that the use of dynamic scaling led to an increase in the number of people identifying the motion as belonging to a child compared to the original adult motion. Our findings suggest that although the dynamic scaling method is not a final solution to translate adult motion into child-like motion, it is nevertheless an intermediate step in the right direction. To better illustrate the original and dynamically scaled motions for the purposes of this paper, we rendered the dynamically scaled motion on an androgynous manikin character.

CCS CONCEPTS

• **Computing methodologies → Animation, Motion capture, Perception**;

KEYWORDS

Perception of motion, Point light displays, Biological motion, Child motion, Markerless motion capture, Algorithm

MiG '17, November 8–10, 2017, Barcelona, Spain

ACM ISBN 978-1-4503-5541-4/17/11…$15.00
https://doi.org/10.1145/3136457.3136460

ACM Reference format:
Yuzhu Dong, Sachin Paryani, Neha Rana, Aishat Aloba, Lisa Anthony, and Eakta Jain. 2017. Adult2Child: Dynamic Scaling Laws to Create Child-Like Motion. In *Proceedings of MiG '17, Barcelona, Spain, November 8–10, 2017,* 10 pages.
https://doi.org/10.1145/3136457.3136460

1 INTRODUCTION

Video games and electronic entertainment have become daily routine for many children. Children under age 11 spend over 1 hour on TV/video and 0.54 hours on computer games per day [Christakis et al. 2004]. Therefore, related industries have increasingly targeted children to meet the expanding demand. Researchers have also found that customizing characters to be similar to the user can increase enjoyment of an experience during game play [Hwa Hsu et al. 2005]. Companies such as Sony Computer Entertainment and Telltale Games [Wikipedia 2017a,b] have developed adventure games such "The Last of Us" and "The Walking Dead" using children as their main characters to attract young customers. To create those characters and bring them to life, animators need to customize their appearance, voice, motions and so on to match their identity as a child. In this paper, we focus on tailoring the motion in a way that looks like it was performed by a real child.

Motion capture (mocap) data is widely used to animate avatars. Compared to animators creating the animation poses from scratch, mocap data has the advantage of portraying realistic motions from real actors [Menache 2000]. Across the existing mocap databases, adult mocap data are more abundant and comprehensive compared to child mocap data[Gross and Shi 2001]. This difference could be because children may have a hard time following the instructions and easily get distracted [Piaget 2015], therefore slowing down the capture process, not to mention the additional cost and effort associated with hiring minors.

Previous works have shown that people are able to identify whether an abstract motion was performed by a child or an adult at levels significantly above chance [Jain et al. 2016]. These findings suggest that if adult mocap data is used to drive a child character, the result would likely not be very compelling. We hypothesized that if adult motion capture data could be stylized to have child-like characteristics, then existing adult mocap databases could be used to drive child characters in games and animations.

There is a large body of literature in motion stylization [Hsu et al. 2005; Ikemoto et al. 2009; Min et al. 2010; Wang et al. 2007; Xia et al. 2015]. However, these approaches either require long training sequences [Hsu et al. 2005] or a large amount of training samples [Xia et al. 2015]. We note that previous work in biomechanics, robotics, and computer graphics have successfully proposed dynamic scaling laws to transfer motion from one character to its scaled version [Hodgins and Pollard 1997; Raibert and Hodgins 1991]. These laws provide a set of scaling rules that scale the height of the actor and the duration of the motion in a way that preserves physics constants such as the value of gravity [Hodgins and Pollard 1997; Raibert and Hodgins 1991]. As a result, a stylization procedure based on these rules should apply to a wide variety of motions, without requiring the collection of exemplar motion sequences from child actors. We created child-like motion from adult mocap data by scaling the time and the displacement values using dynamic scaling laws. We evaluate the effectiveness of this method via a perception study. This evaluation is performed by rendering the dynamically scaled motion as a point light display video, and comparing it to point light display videos of adult motion and child motion. Naïve viewers were asked to identify if the motion was belonging to a child or an adult. We hypothesized that adult motion, when translated to child motion using dynamic scaling laws, would be perceived as child motion. Our results showed that viewers are more inclined to perceive the dynamically scaled adult motion as child motion, though not to the same extent as the original child motion. Our evaluation additionally replicated previous findings by Jain et al. [2016] that viewers are able to distinguish adult motion from child motion at levels significantly above chance. We also show the dynamically scaled motion on an androgynous wood manikin figure.

The contributions of this paper are (a) the application of dynamic scaling laws to create child-like motion from adult motion capture data, and (b) a perception study using point light display videos of child motion, adult motion, and dynamically scaled adult motion that shows that viewers perceive dynamically scaled adult motion as child motion significantly more often than original adult motion, but less so than original child motion.

2 RELATED WORK

We first review the prior work on perception studies in terms of point light displays and rendered characters. We then focus on differences in motion between children and adults. Finally, we discuss different motion translation techniques.

2.1 Perception studies

Researchers have found that people can infer information about actors based on point light display representations (e.g., as shown in Figure 1), which contain only dots representing each joint [Atkinson et al. 2004; Barclay et al. 1978; Jain et al. 2016; Johansson 1973]. Studies have shown that a point light display can convey information regarding the motion type, the actor's identity, gender, and even emotions. Barclay et al. [1978] found that the gender of actors is identifiable from body movements using point light displays. Their study also implied that presentation rate affects gender recognition: non-normal speed will break physics laws in motions and in return alter the perceived naturalness. A study conducted by Atkinson et al. [2004] also showed that viewers can successfully identify actors' emotions at levels greater than chance. They found that participants were better able to identify the emotions when presented with multiple frames of the actors' motion rather than a single frame. Thus in our study, we chose to present the motions as videos rather than static images to give more clues about the motion to viewers. A perception study by Jain et al. [2016] also showed that viewers are in general successful at distinguishing adult actors from child actors when viewing point light display videos of their motion. We use the publicly available data [Jain 2016] from this study for our reference videos and mocap data for adult and child motion; our study validates and extends their results.

Furthermore, previous work has shown that a character's appearance has a significant impact on people's perception [Chaminade et al. 2007; Hodgins et al. 1998; Narang et al. 2017]. People have a higher accuracy rate recognizing themselves from photo-realistic characters than point light displays [Narang et al. 2017]. However, the problem is that the rendered character can mislead viewers due to its appearance. Chaminade et al. [2007] found that, when viewers saw motion on fully fleshed characters, they were more biased toward believing the motion is biological rather than artificial. It is not known if appearance of a child character alone would be sufficient to make the motion seem "naturally" child-like, however. In our case, we removed the appearance entirely by relying on the point light display paradigm so that viewer judgments were made only on motion characteristics. Our approach allows us to isolate "child-like" motion to determine if adult motion can be translated to child characters and still seem naturally child-like.

2.2 Motion differences in children and adults

Previous work has studied differences in the motion characteristics of child motion and adult motion in different contexts. Davis et al. [2001] investigated the differences in walking motion between children and adults by analyzing their gait features. They found that children generally complete strides faster than adults. Their study focused on younger children (ages 3 to 5). Our study focused on children ages 5 to 9 because children in this age range undergo a rapid improvement in motor performance as they develop [Thomas 1980]. Horita et al. [1991] compared the body configurations and joint functions in 6-year-old and adult males for a standing jump motion. By analyzing body segments and joint angles, they found adult performance was much better than children with respect to body motor control. Our study expands on these previous studies by considering an age range from 5 to 9 years, and a wider range of six different actions.

2.3 Motion translation

In previous work, motion translation techniques have been used to change a character's gait [Hsu et al. 2005], conveyed emotion [2015], and body proportions [Gleicher 1998]. Hsu et al. [2005] found that input motion can be transformed into a new style using linear invariant models in real time. For example, a normal walk can be translated into a crouched walk. Furthermore, with an online learning algorithm, relationships between style and motion can be established, which means unlabeled heterogeneous motion can be identified and then translated [Xia et al. 2015]. However, these approaches either require long training sequences [Hsu et al. 2005] or a large amount of training samples [Xia et al. 2015].

Table 1: If the height of an actor is scaled by factor L, its motion attributes will also be scaled according to the table [Hodgins and Pollard 1997].

Quantity	Units	Geom. Scaling	Mass Scaling
basic variables			
length	L	L	
time	T	$L^{1/2}$	
Motion variables			
displacement	L	L	
velocity	LT^{-1}	$L^{1/2}$	
acceleration	LT^{-2}	1	
angular displacement		1	
angular velocity	T^{-1}	$L^{-1/2}$	
angular acceleration	T^{-2}	L^{-1}	

Raibert and Hodgins [1991] and Hodgins and Pollard [1997] applied dynamic scaling laws to computer generated motion of both humans and animals to allow an animator to adapt these motions to humans and animals of different heights. The researchers also pointed out that, if a motion control system is scaled in a certain fashion, motion parameters will also be scaled accordingly. Hodgins et al. [1997] used the scaling laws to modify the controller parameters for a control system that generated adult human motion so that it would generate motion for a child character. They also pointed out that dynamic scaling laws can be applied to motion capture data directly. Inspired by this work, we applied dynamic scaling laws to motion capture data in our paper to create child-like motion from adult motion capture data, with the goal of doing away with the need for new motion capture sessions for child avatar animation.

2.4 Rendering characters

To see the effects of the scaling laws on the motion for the purposes of this paper, after the perception study, we rendered the motion onto an androgynous manikin character. Rendered characters contain richer information about the motion than point-light displays, and are often used in character animations. Hodgins et al.'s [1998] study asserted that subjects are more sensitive to the changes in running motion in a rendered geometric model compared to a stick figure model. Hence, we rendered our motion on a wooden figure model rather than a stick figure in order to amplify and make more visible the characteristics of the motions such as the speed. However, rendering realistic motion that matches motion capture data accurately is challenging, especially for characters with different skeletons. Gleicher et al. [1998] created a space-time constraint solver that adapts the motion while maintaining desirable features of the original motion. We guarded against this problem by modifying the skeleton and the skin mesh of the virtual characters to match the actors. Feng et al. [2014] also proposed a pipeline to tackle the challenges faced when preserving physical constraints of the motion such as foot sliding and foot penetrating. We tackled the foot sliding and foot penetrating problem by modifying the position of the hip joint to ensure the foot always stays right on the ground.

Figure 1: Examples of the test condition for a Jumping Jacks motion shown to the participants. Left: Adult, Center: Child, and Right: Dynamically Scaled Adult.

3 BACKGROUND: DYNAMIC SCALING LAWS

Previous work in biomechanics, robotics, and computer graphics have successfully proposed dynamic scaling laws to transfer motion from one biped character to its scaled version [Hodgins and Pollard 1997; Raibert and Hodgins 1991]. More specifically, when the body of a human character is scaled by a certain factor L uniformly across all dimension, its new motion can be found using dynamic scaling laws (Table 1). The derivation of Table 1 is based on the assumption that the acceleration of gravity for the two characters is constant. According to Newton's law of motion, $a = d/T^2$, the acceleration a can be found from the displacement d an object moves and the time it takes T. For ballistic motions such as jumping, where the acceleration is the acceleration due to gravity g, if the displacement is scaled L times, the time duration must be scaled by $\sqrt{L}$ to maintain g. This change implies that the velocity of the scaled motion can be derived as $v = L/T = \sqrt{L}$. Table 1 summarizes this conversation between the scaled motion and the original motion. We assumed the dynamically scaled adult and original adult are geometrically similar. In other words, the skeleton of a dynamically scaled adult can be generated from scaling an adult skeleton uniformly along all dimensions. In our case, because all the actions were performed in place and they only required the body to move up and down (along y-axis), the lateral (along z-axis) and horizontal (along x-axis) movement can be ignored. Except the hip joint, the position of all other joints can be represented as Euler angles (see Section 5) with respect to their parent joint. According to dynamic scaling laws, the angular displacement stays the same during the scaling procedure. Hence only the position of the hip joint is affected. Therefore, we applied the scaling laws to the following parameters:

(1) The displacement of the hip joint along the y-axis.
(2) The time duration T to complete the action.

4 STIMULI PREPARATION

To prepare the stimuli (point light displays) used in our current study, we used the motion capture data collected by Jain et al. [2016]. This dataset consists of movement data of four adult actors (ages 22 to 32, all male) and four child actors (ages 5 to 9, two female) performing a set of six actions, tracked by the Microsoft Kinect.

Figure 2: A MATLAB plot showing the frontal view of 20 joint positions of an adult actor, connected with lines.

The six actions are "Run as Fast as You Can" ("Run Fast"), "Walk in Place" ("Walk"), "Jump as High as You Can" ("Jump High"), "Fly Like a Bird", "Wave Your Hand" ("Wave"), and "Do 5 Jumping Jacks" ("Jumping Jacks"). For more information on the method used to collect the data, see Jain et al. [2016]; the data is publicly available [Jain 2016]. The actors began each action with a T-pose. To generate our stimuli, we discarded the frames corresponding to the T-pose, as well as the extra frames after the motion was completed. We present the steps we used to apply dynamic scaling laws to the adult motion.

(1) For each adult actor and action, we read in the csv file that contains the world coordinate 3D positions of 20 joints of the actor performing the desired action (Figure 2). We denote the joint positions by ${}^{x}_{w}u^{j}_{i}$ for the jth joint and the ith frame. The left superscript x denotes that this is the x-coordinate, and the left subscript w denotes that the measurements are relative to the world coordinate frame.

(2) The 20 joints can be represented as a tree structure as shown in Figure 3, with the hip center as the root. We created a local coordinate frame that is parallel to the world coordinate frame but has its origin at the parent joint of a given joint j (see Figure 4). The world coordinate 3D position ${}^{x}_{w}u^{j}_{i}$ of a joint j was converted into the local coordinate 3D position ${}^{x}_{l}u^{j}_{i}$ by subtracting the position of its parent joint. We started at each extremity and moved up the tree structure shown in Figure 3.

(3) Figure 4 shows the local coordinate frame of a skeleton from a single actor where the origin is in the parent joint and the axes are parallel to the world coordinate. Because the hip joint does not have a parent joint since it is the root joint, we kept its world coordinates. The x and z displacement of the hip were not scaled because all of the actions were performed in place.

(4) We assume that a child skeleton can be generated by uniformly scaling the adult skeleton. That is, in this work, we do not account for the fact that children have differently proportioned limbs. The joint positions of the dynamically scaled body can be found from equation (1). The scale factor L is the ratio of the average height of a child for ages 5 to 9 years (recorded as 1.22 meters by the World Health Organization's (WHO) Growth Reference [Onis et al. 2007]), and the height of the input adult. The height of the input adult is obtained by computing the distance between the y position of the head ${}^{y}_{w}u^{hip}_{1}$ and the ground which is computed as the y position of whichever the foot is lower ${}^{y}_{w}u^{lowerfoot}_{1}$ in the first frame. This distance is marked in Figure 2.

$$ {}^{x}_{l}u^{j'}_{i} = L \cdot {}^{x}_{l}u^{j}_{i} \tag{1} $$

(5) Because we assumed that the scaled skeleton is geometrically similar to the original skeleton, the relative joint position in local coordinate space does not change. Only the displacement of the hip center joint and the time duration were scaled. For each frame, we computed the vertical displacement of the hip joint from its position in the first frame, shown in equation (2).

$$ \Delta {}^{y}_{l}u^{hip}_{i} = {}^{y}_{l}u^{hip}_{i} - {}^{y}_{l}u^{hip}_{1} \tag{2} $$

The relevant dynamic scaling laws attributes we used in our procedure are shown in Table 1. We multiplied the displacements in the y direction by the scaling factor L. Therefore, the y position of the dynamically scaled motion is computed as the sum of these scaled displacements and the y position in the first frame as shown in equation (3).

$$ {}^{y}_{l}u^{\prime hip}_{i} = {}^{y}_{l}u^{hip}_{1} + L \cdot \Delta {}^{y}_{l}u^{hip}_{i} \tag{3} $$

(6) Since the acceleration due to gravity is constant (in units of $m \cdot s^{-2}$), the duration of the motion must be scaled $\sqrt{L}$ times if the vertical displacement is scaled L times. This scaling process reduces the number of frames, thus keeping the frame rate consistent across the videos. The total number of frames for the dynamically scaled motion was calculated according to equation (4).

$$ T' = \mathrm{round}\left(\sqrt{L} \cdot T\right) \tag{4} $$

Figure 3: A plot demonstrating the tree structure of the actor's skeleton containing 20 joints.

Figure 4: The local coordinate frame which has the origin of each joint in its parent joint and axes parallel to the world coordinate frame.

Figure 5: The local coordinate frame where the y-axis of the local coordinate frame aligns with the parent limb and the x-axis aligns with the projection of the parent limb on x-z plane in its local coordinate frame.

(7) For each frame i' in the dynamically scaled motion, we found the corresponding frame i in the original adult motion with the mapping in equation (5).

$$i = 1 + (i' - 1)\,T/T' \qquad (5)$$

Each frame i that corresponded to a non-integral frame number was linearly interpolated between $\lfloor i \rfloor$ and $\lceil i \rceil$ to calculate the joint position, shown in equation (6).

$${}^{y}_{l}u^{hip}_{i'} = (\lceil i \rceil - i)\,{}^{y}_{l}u^{hip}_{\lfloor i \rfloor} + (i - \lfloor i \rfloor)\,{}^{y}_{l}u^{hip}_{\lceil i \rceil} \qquad (6)$$

(8) Joint positions in local coordinates were converted back to world coordinates by traversing the tree from the root to the end joint. The world coordinate of a joint is the sum of its parent coordinate and the world coordinate of its parent joint.

(9) In order to minimize bias from the actor's size, we scaled the stimuli to a canonical height using the method proposed by Jain et al. [2016].

5 RENDERING THE MOTION ON A SKINNED CHARACTER

In order to better evaluate our result, we rendered the mocap data on realistic characters. The process included the following series of steps.

(1) The noise in the motion capture data resulted in slight changes in the length of the character's limbs. A one-dimensional median filter of window size 6 was applied on each degree of freedom to smooth the motion. This window size corresponds to 0.2 seconds because the motion was captured at 30 frames per second.

(2) Because the noise in the data resulted in a slightly different limb length in each frame, the frame-wise limb lengths were averaged to obtain the limb length we used in all further computations. We denote the limb length associated with joint j as r^j. Because the joint position has already been converted to local coordinates relative to the parent joint (see Section 3, point (2)), we can compute the limb length as in equation (7).

$$r^j = \frac{1}{T}\sum_{i=1}^{T}\sqrt{({}^{x}_{l}u^j_i)^2 + ({}^{y}_{l}u^j_i)^2 + ({}^{z}_{l}u^j_i)^2} \qquad (7)$$

(3) Every joint was re-parametrized using three Euler angles using intrinsic rotations in x-z-y ordering $({}^{R_x}_{l}u^j_i, {}^{R_y}_{l}u^j_i, {}^{R_z}_{l}u^j_i)$ relative to a local coordinate frame where the y-axis is parallel to the parent limb (Figure 5). For the root, that is, the Hip Center, the local coordinate frame is kept parallel to the world coordinate frame. For all subsequent limbs, the local coordinate is aligned to the parent limb. Figure 3 shows the Hip Center.

(4) We then rotated the joints $({}^{x}_{l}u^j_i, {}^{y}_{l}u^j_i, {}^{z}_{l}u^j_i)$ in the second level of the tree hierarchy with respect to the parent joint coordinate. All the angles were represented using the right handed coordinate frame: the y-axis of the parent joint coordinate is aligned with its parent limb and the x-axis is aligned with the projection of the parent limb on the x-z plane in its local coordinate. ${}^{y}_{l}u^j_i$ was set to zero because we do not have enough information to compute its value. We computed the joint angle iteratively using the same process described above. Algorithm 1 demonstrates this process.

(5) After we modified the limb length, we wrote a MATLAB script to convert the mocap data from a csv file to a bvh file.

(6) After the limb length is modified, the motion may violate kinematics constraints from the environment. For example, the changes in the leg limb length can cause the foot to detach from the ground. To resolve this problem, we first annotated the frame number i where either one of the two feet is supposed to touch the ground. Consider one of these frames, which we denote as frame number i. The new position of the foot joint in frame i in the world coordinate after limb length is modified is denoted as $({}^{x}_{w}u^{foot}_i, {}^{y}_{w}u^{foot}_i, {}^{z}_{w}u^{foot}_i)$. We computed the vertical distance between ground position and the foot position as in equation (8).

ALGORITHM 1: Joint Angle Computation

Input: Joint position in local coordinate $\left({}^{x}_{l}u^{j}_{i}, {}^{y}_{l}u^{j}_{i}, {}^{z}_{l}u^{j}_{i} \right)$

Output: Joint rotation in local coordinate $\left({}^{Rx}_{l}u^{j}_{i}, {}^{Ry}_{l}u^{j}_{i}, {}^{Rz}_{l}u^{j}_{i} \right)$

for each frame i do
  for each branch in the tree structure do
    for each joint j from root to leaf do
      ${}^{Ry}_{l}u^{j}_{i} = 0$
      if joint=root node then
        ${}^{Rz}_{l}u^{j}_{i} = \arctan\left({}^{z}_{l}u^{j}_{i}, {}^{y}_{l}u^{j}_{i} \right)$
        ${}^{Rx}_{l}u^{j}_{i} = \arctan\left({}^{y}_{l}u^{j}_{i}; \sqrt{ {}^{y}_{l}u^{j\,2}_{i} + {}^{z}_{l}u^{j\,2}_{i} } \right)$
        $R = rotZ\left({}^{Rz}_{l}u^{j}_{i} \right) \; rotX\left({}^{Rx}_{l}u^{j}_{i} \right)$
      else
        $\left({}^{x}_{p}u^{j}_{i}, {}^{y}_{p}u^{j}_{i}, {}^{z}_{p}u^{j}_{i} \right) = \left[{}^{x}_{l}u^{j}_{i}, {}^{y}_{l}u^{j}_{i}, {}^{z}_{l}u^{j}_{i} \right] R$
        ${}^{Rz}_{l}u^{j} = \arctan\left({}^{z}_{p}u^{j}_{i}, {}^{y}_{p}u^{j}_{i} \right)$
        ${}^{Rx}_{l}u^{j} = \arctan\left({}^{y}_{p}u^{j}_{i}; \sqrt{ {}^{y}_{p}u^{j\,2}_{i} + {}^{z}_{p}u^{j\,2}_{i} } \right)$
        $R = R \; rotZ\left({}^{Rz}_{l}u^{j}_{i} \right) \; rotX\left({}^{Rx}_{l}u^{j}_{i} \right)$
      end
    end
  end
end

$$dis_y = {}^{y}_{w}u^{foot}_{i} - {}^{y}_{w}u^{lowerfoot}_{1} \quad (8)$$

where ${}^{y}_{w}u^{lowerfoot}_{1}$ is they position of whichever foot was lower in the rst frame. To avoid the foot sliding issue, we computed the horizontal and lateral displacement (see equation 9) of the foot position within consecutive frames when they are supposed to touch the ground. We assumed the foot touches the ground in both frames i and $i+1$.

$$dis_x = {}^{x}_{w}u^{foot}_{i+1} - {}^{x}_{w}u^{foot}_{i} \quad (9)$$

$$dis_z = {}^{z}_{w}u^{foot}_{i+1} - {}^{z}_{w}u^{foot}_{i} \quad (10)$$

Then we adjusted the position of the skeleton to place the foot joint in place by subtracting the distance from the hip center, as follows.

$${}^{y}_{w}u^{HipCenter}_{i} = {}^{y}_{w}u^{HipCenter}_{i} - dis_y \quad (11)$$

$${}^{x}_{w}u^{HipCenter}_{i} = {}^{x}_{w}u^{HipCenter}_{i} - dis_x \quad (12)$$

$${}^{z}_{w}u^{HipCenter}_{i} = {}^{z}_{w}u^{HipCenter}_{i} - dis_z \quad (13)$$

(7) We modi ed the geometry and the skeleton to match our actors, and exported the angle of each joint as an .animExport le in Maya. Then we imported the exported le onto the skeleton of the virtual character.

We used the rendered videos to interpret the results of our perception study in closer detail. The rendered results helped us to observe characteristics of the dynamically scaled motion such as speed and coordination. The screenshots of the rendered action for the Jumping Jacks action are shown in Figure 13.

6 EXPERIMENT DESIGN

We evaluated the e ectiveness of the procedure in Section 4 in transforming an input adult mocap sequence to appear more child-like. This evaluation was performed by conducting a point light display perception study. The data collection procedure was part of a protocol approved by our Institutional Review Board (IRB). A total of 24 participants (13 female, age range 21 to 29 years, mean=23.96, standard deviation (SD)=2.9) who completed the survey either participated for extra credit in a class or participated voluntarily.

To test our hypothesis (that adult motion dynamically scaled to child motion could be recognized as child motion), we used a within-subjects design. Every participant watched all three types of videos (adult, child, and dynamically scaled adult) and all six actions (Run Fast, Walk, Jump High, Fly like a Bird, Wave, and Jumping Jacks). The experiment was an online survey. We had 72 videos in total (6 motions 4 child actors + 6 motions 4 adult actors + 6 motions 4 adult actors, dynamically scaled). To account for possible fatigue of the participants, we created 4 surveys. Each survey contained a total of 54 stimuli videos (6 motions 3 child actors + 6 motions 3 adult actors + 6 motions 3 adult actors, dynamically scaled). The actors were counterbalanced across all the possible subsets of actors within an actor type. The same adult and child actor were always paired across di erent surveys. To minimize ordering e ects, the videos were presented to the participants in random order.

Each video was followed by three questions. The rst one was a two-alternative forced choice question, Does this motion belong to a Child or an Adult? The response was recorded via a radio button that allowed only one option (Child or Adult) to be selected. The second one was a 7-point unipolar Likert scale question for the participants to indicate their own con dence in their response. Additionally, participants were asked to enter a free form text answer to the question, What is the action being performed? This question was to ascertain that they had in fact played the videos and identi ed the actions correctly. The videos were rendered at 30 fps. We used the child and adult point light display videos from Jain et al. [2016] and supplemented them with dynamically scaled adult point light display videos that we generated.

7 ANALYSIS AND RESULTS

We analyzed the survey responses to understand (a) how often participants correctly identi ed motion as belonging to a child or an adult, (b) how often they attribute dynamically scaled adult motion as belonging to a child, and (c) how attribution to child or adult is a ected by the action being performed.

7.1 Validation of previous work

Jain et al. [2016] reported the percentage correctness of their survey participants in identifying motions as belonging to a child or an adult. We report our own data on this measure to compare to theirs. In our survey, for each actor type/action pair they saw, each participant answered the survey question, Does the motion belong to a child or an adult? The number of child responses were tabulated per actor type (child, adult, dynamically scaled adult) and also per action type. We rst computed the distribution of overall correctness of adult videos and child videos across all the participants (Figure 6). Because there is no correct option for dynamically

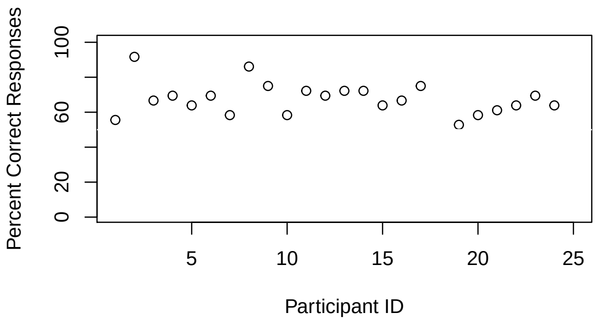

Figure 6: The distribution of overall correctness for each participant. All participants except one have the overall correctness above chance (50%), except one who is thus excluded from remaining analysis.

scaled videos, we did not take this actor type into account when calculating the overall percentage correctness. For each participant, we computed the number of responses that match the actual actor type across all 36 videos which include both adult videos and child videos. We then divided the number of correct responses by 36 to get the overall percentage of correctness. We removed one participant whose overall percentage of correctness was below chance.

To analyze the data obtained, we used the statistical analysis tool, R. First, we ran a Shapiro-Wilk test, which indicated that the distributions of percentage correctness for child and adult videos (Figure 7) were normal ($p = 0{:}8151 > 0{:}05, p = 0{:}8455 > 0{:}05$). We ran a one-tailed t-test for adult videos and child videos, respectively, to examine if the percentage of correctness is significantly above chance. For adult videos, the percentage of correctness (mean=68.36%, SD=12.47%) was significantly above chance (t = 7.0592, df = 22, $p < 0{:}0001$). Similarly for the child videos, the percentage of correctness (mean=66.91%, SD=14.06%) was significantly above chance (t = 5.7664, df = 22, $p < 0{:}0001$). These results confirmed the findings from Jain et al. [2016], that naïve viewers are able to attribute motion rendered in point-light displays correctly to children or adults at levels above chance.

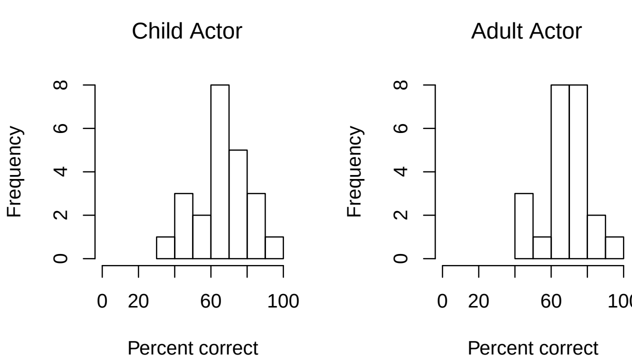

Figure 7: Histogram showing the distributions of the percentage of correctness for child videos and adult videos.

7.2 Results by actor type

To understand how well the dynamic scaling method works in transforming adult motion into child-like motion, we examined how often survey participants identified motion as belonging to a child for each actor type they saw. To do so, we define the value "child response" as the response to the question "Does this motion belong to a child or an adult?". Child response is set to 1 if the response is 'child', and 0 otherwise. We define the number of child responses per participant as the sum of child response for each actor type, for each participant. We ran a Shapiro-Wilk test and found that the distributions of the number of child responses for all videos for each actor type (child, adult, and dynamically scaled (DS)) were normal ($p = 0{:}8151 > 0{:}05$, $p = 0{:}8455 > 0{:}05$, and $p = 0{:}1609 > 0{:}05$, respectively). We computed the means and standard deviations for each actor type (Figure 8, 9a).

We ran a two-way repeated measures ANOVA on *number of child responses* with within-subjects factors of *actor type* (Child, Adult, DS), and *action* ("Fly Like a Bird", "Jump High", "Jumping Jacks", "Run Fast", "Walk", "Wave") and found a significant main effect by actor type ($F_{2,384}$=57.57, p <0.0001). A Tukey post-hoc test showed that the following pairs differ: Adult vs. Child (p <0.001), Child vs. DS (p <0.001), and Adult vs. DS (p=0.019<0.05) (see Figure 10).

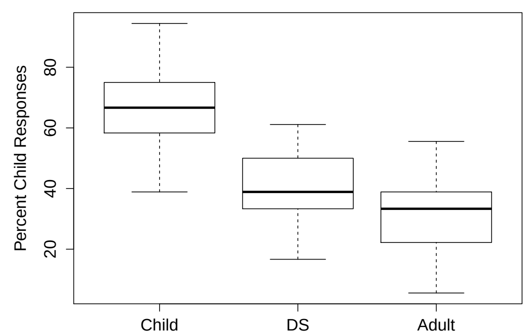

Figure 8: Figure contains boxplot of the distribution of number of child responses. The bold bar in the middle shows the median value.

(a)

Actor Type	Mean	SD
Child	12.04	2.53
Adult	5.70	2.24
DS	7.35	2.17

(b)

Action	Mean	SD
Fly like a bird	4.26	1.76
Jump high	4.57	1.56
Jumping jacks	3.22	1.54
Run fast	4.78	1.54
Walk	3.96	2.12
Wave	4.30	2.10

Figure 9: (a) Means and standard deviations of number of child responses for each actor type. Every participant watched 18 videos of each actor type. The maximal value of number of child response is 18. (b) Means and standard deviations of number of child responses for each action. Every participant watched 9 videos of each action. The maximal value of number of child response is 9.

The number of child responses for actor type Child (mean = 12.04, SD = 2.53) is greater than that for actor type Adult (mean = 5.70, SD = 2.24), which again shows that viewers are able to distinguish child motion from adult motion. For the actor type Dynamically Scaled, the number of child responses (mean = 7.35, SD = 2.17) is greater than that of Adult (mean = 5.70, SD = 2.24), but less than that of Child (mean = 12.04, SD = 2.53). This finding shows that viewers were more likely to perceive the dynamically scaled motion as belonging to a child than they were for the original adult motion. At the same time, viewers did not perceive dynamically scaled adult motion as belonging to a child as often as the original child motion. This ordering indicates that, although the dynamic scaling method is not a final solution to translate adult motion into child-like motion, it is nevertheless a step in the right direction.

7.3 Results by action

We also examined how the identification of motion as belonging to a child was affected by the action being performed. The same two-way repeated measures ANOVA discussed in Section 7.2 showed a significant main effect of action ($F_{5,384}$=3.21, p=0.0075 <0.01).

Figure 11 shows the average number of child responses for all six actions represented in our data set. A Tukey post-hoc test found that the following action pairs differ: "Jumping Jacks" vs. "Run Fast" ($p = 0.004 < 0.01$), and "Jumping Jacks" vs. "Jump High" ($p = 0.023 < 0.05$). Considering the action "Jumping Jacks", viewers are more likely to identify the motion as belonging to an adult regardless of actor type than for the action "Run Fast". One possible reason, based on the results of Jain et al. [2016], could be that "Jumping Jacks" looked generally coordinated, and it is possible that viewers associate better coordination with adult motion. For the "Run Fast" action, both adult and child actors performed them very fast. The overall number of child responses for this action was dominated by responses for child actors (21.82% of child responses belonged to adult actors, 50.91% of child responses belonged to child actors, and 27.27% of child responses belonged to DS actors). We think it is because generally children are less balanced than adults [Schaefer et al. 2008], and the "Run Fast" action prompts the actor's motion to be fast, this child actors may be more likely to lose balance when performing the action in place. The same trend can be found in the action "Jump High". Two of the four child actors did more than one jump in their action, which could have increased the number of child responses for those videos (e.g., for actor 290 who did 4 jumps, the percentage of child responses is 73.91%; for actor 337 who did 3 jumps, the percentage of child responses is 60.87%; for actor 723 who did 1 jump, the percentage of child responses is 52.17%; for actor 644 who did 1 jump, the percentage of child responses is 39.13%,). A longer video helps viewers to infer more information about the actor. However, if we cut off the videos after one jump, we were concerned that the naturalness of the motion will be compromised.

8 DISCUSSION AND FUTURE WORK

In this paper, we created child-like motion from adult mocap data using dynamic scaling laws. We conducted a perception study to evaluate the effectiveness of this method.

We found significant differences in the number of child responses by actor type and action. Viewers were more likely to attribute the dynamically scaled motion to a child than they were for the original adult motion. Yet, the dynamic scaling method does not completely convince the viewers that they are looking at child motion: viewers did not perceive dynamically scaled adult motion as belonging to a child as often as they did for the original child motion. In our survey, participants gave a confidence score on a scale of 1 to 7 for each of their responses to the videos. We analyzed these confidence scores by actor type. A one-way repeated measures ANOVA on *confidence score* with a within-subjects factor of *actor type* (Child, Adult, Dynamically Scaled Adult) found no significant difference

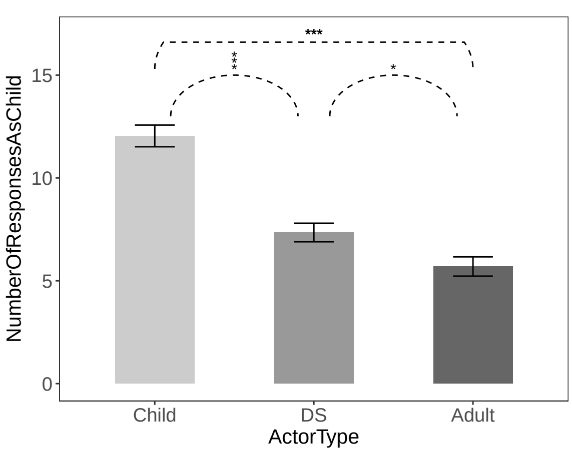

Figure 10: Bar chart showing the mean and standard error of number of child responses for all three actor types. The error bars represent the standard error. * indicates significance at the $p < 0.05$ level, whereas * indicates significance at the $p < 0.001$ level.**

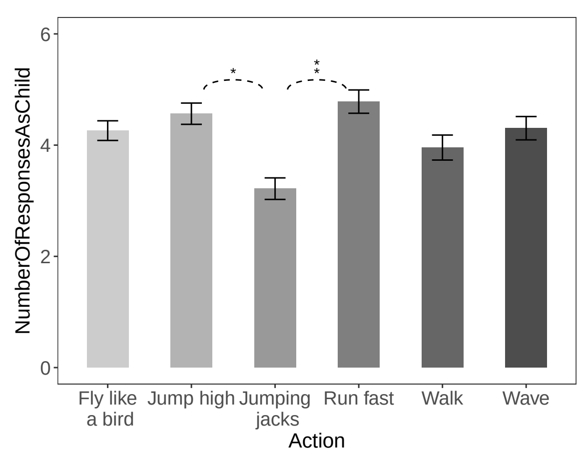

Figure 11: Bar chart showing the mean and standard error of child responses for all six actions. The error bars represent the standard error. * indicates significance at the $p < 0.05$ level, whereas ** indicates significance at the $p < 0.01$ level.

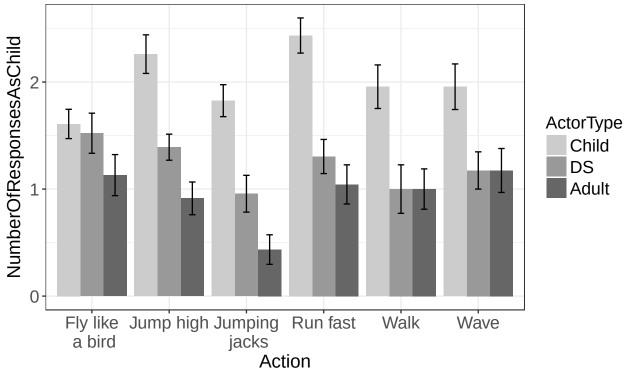

Figure 12: Bar chart showing the mean and standard error of number of child responses per action for all three actor types. The error bar represent standard error.

by actor type ($F_{2,1217}$ = 0.52, *n.s.*). This finding shows that viewers are equally confident about their choices for all actor types, thus making our results comparable among the actor types. Thus, we can conclude that the differences by actor type are reliable.

For action, viewers were generally more likely to attribute the action "Jumping Jacks" to an adult and the action "Run Fast" and "Jump High" to a child, regardless of the actual actor type of the video. This finding indicates that the actions that are included in a dataset may have an effect on viewers' perception. An extreme example of this effect might be the inclusion of the action "Crawling": viewers may be primed to perceive such an action as belonging to a child no matter what the other motion characteristics are. Future work could investigate more actions in a wider range of contexts.

We did not find an interaction effect between actor type and action. However, Figure 12 illustrates some trends of interest. For the "Jumping Jacks" action, viewers are generally more likely to attribute dynamically scaled adult videos as being performed by a child, and less likely to attribute the original adult videos as being performed by a child. However, for the "Walk" and "Wave" actions, viewers are equally likely to attribute dynamically scaled adult videos and the original adult videos as being performed by a child. To better understand how the action being performed affects viewers' perception, we rendered the child, adult, and dynamically scaled adult videos for all 6 actions on wooden androgynous figures (Figure 13 and supplemental video). From watching the rendered videos, we observed that for the action "Jumping Jacks", it is evident that the adult actors were more coordinated than the child actors, but the adults moved slower than the children. For dynamically scaled adults, we observed that the coordination level seemed similar to that of adults, but the dynamically scaled wooden figures moved as fast as child figures. However, we observed that for the "Walk" and "Wave" actions, dynamically scaled adults moved similarly to the original adults. Taken together, these insights imply that viewers may perceive faster actions as belonging to children. In order to adapt the motion to child characters, the dynamic scaling procedure we used (Table 1) shortened the time duration of completing one action and in turn increased the speed. Therefore, the dynamic scaling procedure seems to amplify a motion characteristic that viewers use as a cue to distinguish child from adult motion. It is worth noting that previous work in physiological gait analysis supports the idea that children move faster than adults [Davis 2001]. Future work could extend this study by recruiting more actors to investigate these trends.

It is worth noting that participants in our study may not have interacted with children often enough to properly identify their motion. Future work can consider recruiting participants who have interacted more often with children such as parents and teachers. Also, despite our decreasing the total number of videos that survey participants saw from 72 to 54, some participants still complained of fatigue. The survey system estimated 49 minutes as the total time to complete the survey. This fatigue may have affected how accurate responses were later in the survey. We controlled for this issue by randomly ordering the videos that each participant saw. Another limitation is that dynamic scaling laws only changed the displacement and time duration of the motion. Because we are dealing with motion capture data, the scaling in mechanical parameters such as force, mass, or stiffness is not applied and hence, any such relevant changes between child and adult motion are not captured. Also, dynamic scaling laws are not sufficient to alter the coordination and joint angle in the motion which could be an important cue for viewers to tell apart children from adults. Because we assumed that the skeleton of adult and child are geometrically similar, the variations in limb ratio were not captured. Future work could investigate an algorithm to retarget the motion to child characters that have different limb ratios and mass distributions, as well as different coordination and motion control abilities. In conclusion, this work takes an important first step in translating adult motion into child-like motion, and can be of use in future work in games, animation, and virtual reality.

9 CONCLUSION

In this paper, we considered the problem of generating motion for child characters in games and animations by leveraging existing databases of adult motion capture data. We showed that, by simply changing the time and the displacement value of the motion using dynamic scaling laws, we could create child-like motion from input adult mocap data. We presented a discussion of the dynamic scaling procedure, and algorithmic details for how to implement it. Finally, we tested if naïve viewers perceive dynamically scaled adult motion as child motion through a perception study. Our results showed that, although viewers judged dynamically scaled adult motion as belonging to a child significantly more often than the original adult motion, they still attributed it to a child less often than the original child motion. Because the dynamic scaling procedure is simple to implement, does not require a training or exemplar database, and achieves some success in convincing viewers, this method can be used to create child motion from adult motion for a variety of relevant applications, such as games, avatars in online education, and virtual characters such as museum tour guides.

ACKNOWLEDGMENTS

The authors wish to thank Julia Woodward, Annie Luc, Alex Shaw, Isabella Cuba, Amanda Castonguay for helping with data collection.