Is the motion of a child perceivably different from the motion of an adult?

EAKTA JAIN and LISA ANTHONY and AISHAT ALOBA, University of Florida
AMANDA CASTONGUAY, University of Southern Maine
ISABELLA CUBA, Vassar College
ALEX SHAW and JULIA WOODWARD, University of Florida

Artists and animators have observed that children's movements are quite different from adults performing the same action. Previous computer graphics research on human motion has primarily focused on adult motion. There are open questions as to how different child motion actually is, and whether the differences will actually impact animation and interaction. We report the first explicit study of the perception of child motion (ages 5 to 9 years old), compared to analogous adult motion. We used markerless motion capture to collect an exploratory corpus of child and adult motion, and conducted a perceptual study with point light displays to discover whether naive viewers could identify a motion as belonging to a child or an adult. We find that people are generally successful at this task. This work has implications for creating more engaging and realistic avatars for games, online social media, and animated videos and movies.

CCS Concepts: •**Computing methodologies** → **Animation;** *Motion capture;* **Perception;**

Additional Key Words and Phrases: Perception of motion, Point light displays, Biological motion, Child motion, Markerless motion capture

1. INTRODUCTION

Artists and animators who draw child characters have observed that children's poses, gestures, and movements are quite different from adults performing the same action [AnimationAddicts 2013]. There are several possible factors that cause these differences. For example, researchers have studied the biomechanical differences between child and adult body proportions and body mass distributions [Huelke 1998]. Child development experts have studied the development of the neuro-muscular control system [Thelen 1995; Assaiante 1998]. Young children generally have more energy than teens and adults, and that might manifest itself in the ways they move around [Nader et al. 2008]. Regardless of the reason, there is a general consensus based on conventional wisdom that child motion is quite different from adult motion.

In spite of the long history of research on human motion, especially in the context of character animation for movies and games, these efforts have been focused primarily on adult motion. Our literature review led us to previous systematic investigations of child skeletal and neuro-muscular development [Thelen 1995; Huelke 1998; Assaiante 1998], and motion capture of children for studying motor impairments [Sandlund et al. 2009; Chia et al. 2013; Rosengren et al. 2009], but no significant

Author's address: Eakta Jain, Lisa Anthony, Aishat Aloba, Alex Shaw, Julia Woodward, Department of CISE, University of Florida, P.O. Box 116120, Gainesville, FL 32611; email: ejain@cise.ufl.edu; Amanda Castonguay, University of Southern Maine, P.O. Box 9300, Portland, ME 04104 (Work done as a summer intern at the University of Florida); Isabella Cuba, Vassar College, 124 Raymond Avenue, Poughkeepsie, NY 12604 (Work done as a summer intern at the University of Florida).

 1544-3558/2016/XXXX-ARTXXXX $15.00
DOI: http://dx.doi.org/10.1145/2947616

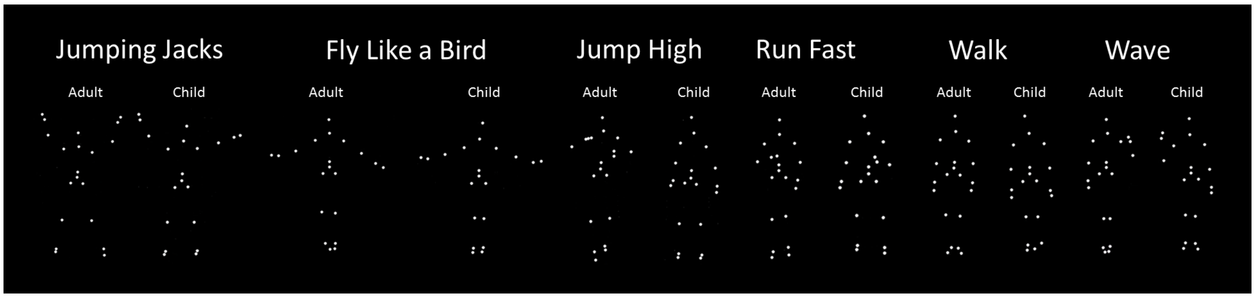


Fig. 1: Sample frames from our point light displays of child and adult motion we investigated in this study.

study of children's movements from a perceptual perspective. As games, movies, social media, and online education are increasingly geared towards younger audiences, it becomes important to study child motion for the purposes of character animation. The particular question that we consider in this work is whether child motion actually is perceivably different from adults. If the answer is yes, then it would open the door to several interesting questions, ranging from how different child motion actually is, to how such differences could be quantified, and whether such differences will actually impact animation and interaction. Answers to these questions would influence the development of more engaging and realistic avatars for children.

We present the results of the first explicit study of the perception of child motion (ages 5 to 9 years old) and comparison to analogous adult motion. We first collect a corpus of actions performed by child and adult actors, using markerless motion capture via the Microsoft Kinect[1]. Our corpus contains motion from eight actors (four children, four adults) performing six dynamic actions: Fly Like a Bird, Jump as High as You Can, Do Five Jumping Jacks, Run as Fast as You Can, Walk in Place, and Wave Your Hand (Figure 1). These actions were selected to span a range of dynamic motions, and include common motions typically studied in similar perceptual experiments [Johansson 1973; Dittrich et al. 1996; Cutting 1978]. We then transformed the captured motion into point light displays, scaling the rendered motion to a standard height to control for natural height differences between children and adults. We conducted a perception study to discover whether independent participants could identify a point light display as belonging to a child or an adult.

Among all the open questions in the study of child movement from the perspective of animation and interaction, we believe that the perception question is important to address first: if the differences are not perceivable, then there is no need to build special models for animating avatars to have child-like movements. While action types, gender, emotion, and more have been investigated in the perceptual literature [Johansson 1973; Cutting 1978; Barclay et al. 1978; Dittrich et al. 1996], the question of the perception of child motion has not been studied, possibly because generating stimuli by motion capturing children is challenging. Recent advances in markerless motion capture technology allowed us to record children as young as five years old, and to prepare a database of motions from multiple child and adult actors that enables researchers to conduct controlled experiments.

Because this is the first study of its kind, we choose to record child and adult actors performing actions that are comparable to the extent that both sets of actors receive the same instructions, but they are otherwise allowed to interpret the instructions naturally. For example, when instructed to "Run as fast as you can", the child does run as fast as he can, whereas the adult runs at a less rapid pace. We do not allow the experimenter to prod the adult in such a case. We allow these differences to

[1] https://www.microsoft.com/en-us/download/details.aspx?id=40278

remain in our corpus because our motivation is to study child and adult movements for the purpose of generating more compelling avatars, and for enabling interaction and recognition systems to work with different types of users. A more strictly controlled study might have asked both sets of actors to run on a treadmill with a pre-set speed, and for a predetermined duration, to factor out differences in task interpretation in favor of more precisely controlled motion data. However, for our purposes, this design decision would also have washed out an interesting observation that affects the design of interactive experiences: children are quite enthusiastic in their movements, whereas adults are more reserved.

The main contributions of this paper are (a) the application of the point light display paradigm towards studying the movement of young children and adults performing the same actions, and (b) the finding that naive viewers can identify a motion as belonging to a child actor or an adult actor at better-than-chance levels in a two-alternative forced choice task. We release the corpus of point light videos as a resource for the community[2].

2. BACKGROUND

We first briefly review related work on the motion capture of children and how it has been used. We then review work on the perception of human motion and the ability of viewers to identify characteristics of human actors from moving point light displays. (For a detailed survey, see Blake and Shiffra [2007].) We also review relevant results from the perception of animated characters, since this work can inform the design of animated avatars for children.

2.1 Study of child motion

Motion capture of children in the biomechanics domain has been done extensively, especially with an aim to compare typically developing children to the motion of children with motor impairments. For example, Sandlund et al. [2009] used motion capture of children and teens (ages 6 to 16) with sensorimotor disorders to detect whether or not rehabilitation interventions were effective. Chia et al. [2013] and Rosengren et al. [2009], in two similar studies, used motion capture technology to compare the running gait of boys ages 7 to 10, and children about age 7, respectively, half of whom had developmental condition disorder (DCD), and characterized these differences quantitatively. Delp et al. [2007] created OpenSim, software to allow the simulation of dynamic muscle forces during movement, which were previously unavailable in simulations; they proved the efficacy of their software by simulating the walking of a 12-year-old child with cerebral palsy. Researchers have also found that motion capture on children can be effective when used to quantify the characteristics of social behaviors versus those in children with developmental disorders [Sakuma et al. 2012].

Some work has compared adult and child motion, but this work has used much younger children (e.g., infants and toddlers up to 3 to 5 years of age [Ivanenko et al. 2013; Davis 2001]. Differences that have been identified between adult and young children's motion centered on muscle activation and control [Ivanenko et al. 2013], and stride frequency or configuration [Davis 2001].

2.2 Perception of human motion

Gunnar Johansson famously showed that viewers could recognize human actions even with extremely sparse information in the form of point lights on the joints of actors [Johansson 1973]. Since then, researchers have found that point light representations can be used by observers to glean a great deal of information about the action and the participant. We now briefly review the various characteristics that viewers have been able to infer about actors by watching these point light displays in prior work.

[2] http://jainlab.cise.ufl.edu/pose-perception.html

Cutting et al. [1978] found that observers were able to distinguish male and female motion based on point light displays, and offered theories on what allowed viewers to make this distinction. Barclay et al. [1978] also examined the perception of gender from point light displays, and found that there is a threshold of time when viewing the motion before which a viewer's ability to choose gender was no better than chance. This finding implies that the ability to distinguish gender is a property of the motion of the points rather than just their position. Pollick et al. [2002] noted that among several factors, the velocity of an action particularly affects a viewer's ability to classify a point light display actor as male or female. In a more recent study, Brooks et al. [2008] found a correlation between the orientation of a point light display (that is, the direction the actor is facing) and the perceived gender of the actor.

Another property that has been examined using point light displays is the perception of emotion. Dittrich et al. [1996] found that participants were able to recognize emotions such as fear, grief, and joy from point light display videos of dancers, but that the ability to recognize these emotions was reduced when the motion was inverted. A study by Atkinson et al. [2004] examined viewers' ability to distinguish emotions such as happiness, sadness, and fear from point light display videos as well as full light display videos of the actions. The results showed that, while correct classification of emotions was higher for full light display videos, the correct rate of responses for the point light videos was well above chance.

A number of studies have further shown that viewers are able to represent point light motion as belonging to specific people, particularly themselves. Cutting and Kozlowski [1977] showed that viewers were able to identify motion belonging to their friends via a point light display of them walking at a much higher rate than chance. Beardsworth and Buckner [1981] built on this work to show that viewers were able to recognize point light videos of themselves at an even higher rate than those of friends. Loula et al. [2005] asked participants to classify point light motion as belonging to themselves, a friend, or a stranger. They found that viewers were most accurate at identifying their own motion, followed by that of friends and then by that of strangers. Wellerdiek et al. [2013] reported that users were able to identify their own motion by observing point light displays as well as by viewing virtual avatars performing their motion.

Studies of motion using point light displays have not been limited to adult viewers. Pavlova et al. [2001] showed that children (ages ranging from 3 to 5 years) can recognize biological motions of humans and animals, but they were unable to recognize these same motions when shown a non-animated series of consecutive frames from the same point light video. A study by Golinkoff et al. [2002] also showed that 3-year-old children were able to appropriately classify human actions by viewing point light videos of motion, and Klin et al. [2009] compared preference for biological motion in toddlers (age 2) with and without autism spectrum disorder. These same displays have been used to demonstrate a preference for biological motion even in infants [Fox and McDaniel 1982]. However, to our knowledge, no work has been conducted to study the perception of point displays of *children's* motion (that is, with child actors) to see if viewers are able to recognize child motions as well as adult motion. Because of the long history of the point light display paradigm to study human motion, we use point light displays for our experiment to investigate the perception of motion recorded from a child actor relative to an adult actor.

2.3 Perception of animated characters

Previous literature has shown that the role of appearance on the perception of human motion cannot be neglected. It has been reported that people are more biased towards declaring a motion as biological (versus synthetic) when it is rendered with a fully fleshed character model rather than a sparse representation [Chaminade et al. 2007; Stadler et al. 2012; Neri et al. 1998; Beardsworth and Buckner

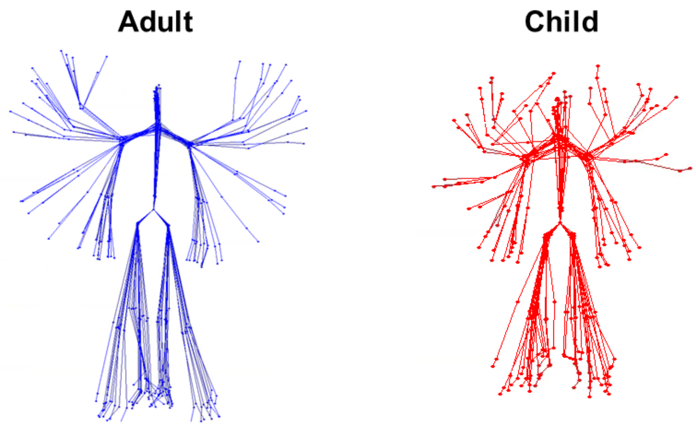


Fig. 2: Motion data (3D joint positions) rendered from the frontal view on every tenth frame for jumping jacks for an adult actor and a child actor (prior to scaling to canonical height). Both figures have been pinned at the pelvis joint. Note how the child actor's elbows bend more, and the overall motion is less coordinated.

1981], and people are more sensitive to variations in dynamic actions such as running for fully fleshed characters [Hodgins et al. 1998]. However, body representation appears not to make a difference to the perception of emotion in actions performed by animated characters [McDonnell et al. 2008].

When considered in the context of these same questions studied using point light displays, these results imply that if people can see an effect with point light displays, they will be able to do so with fully rendered models, but not vice versa. This claim is supported by Hodgins et al.'s study on the effectiveness of various models of displaying motion for perception studies [Hodgins et al. 1998]. In that study, the researchers found that viewers performed better on perception tasks with highly rendered models resembling animated characters than they did with less complex models that resembled stick figures. Harrison et al. [2004] were similarly motivated to study the perception of length changes in animated characters by constructing an experiment around stick figure characters. Because our goal is to verify our hypothesis that people can make a distinction between child and adult motion, regardless of avatar appearance (e.g., child avatars or adult avatars), we start with the point light display paradigm. We hypothesize that if our results reveal an effect with these appearance-impoverished stimuli, then the same effect will be observable on fully fleshed characters as well.

3. STIMULI PREPARATION

To capture the stimuli for our dataset, we solicited paid volunteers to perform body movements and motions in our laboratory. The data collection procedure was part of a protocol approved by our Institutional Review Board (IRB), including the participation of the children.[3] Although the volunteers were not trained actors, we use the term "actors" throughout this paper to distinguish them from participants in the perception study. Data was recorded by a tool we developed using the Microsoft Kinect 1.0 hardware and 1.8 software SDK to track the actors as they completed a series of motions. We used movement data from four adults (ages 22 to 32, all male) and four children (ages 5 to 9, two female) for our actors. We balanced the demographics of the child actors between gender and age: we have two

[3]As is standard with using children as research participants, written parental informed consent was obtained as well as verbal child assent to the procedure.

females, ages 5 and 9, and two males, ages 6 and 8. Each actor performed one repetition of a set of six actions we chose as the stimuli for this experiment. We illustrate the motion data in Figure 2. This visualization has been pelvis-aligned to better illustrate the differences in joint angles and velocities between children and adults.

To increase consistency across actors, we asked them to "count down" to the start of their action by beginning in a T-pose (standing with arms outstretched and feet shoulder length apart, forming a "T" with their body) and slowly dropping their arms as they counted from "3" to "1" as an experimenter began the recording. At "1", they could begin their action. They were asked to return to a standing position with arms at their sides once they were done executing the action. At this point, the experimenter ended the recording of that action. For all actions, we allowed actors to set their own pace and duration. The only instances in which the experimenter asked the actor to redo the action were if an actor performed any given motion for an extremely short duration (e.g., only walking for one or two steps). In this case, they were asked to repeat the action "for a little longer". We targeted five cycles for most actions, except for "Jump High" (one cycle). Note that for "Run Fast" and "Walk in Place", five cycles is equal to ten steps.

The set of six actions are as follows (the exact wording of the prompt given by the experimenter is shown in parentheses):

(1) Wave Your Hand ("the next action is to wave your hand"),
(2) Walk in Place ("the next action is to walk in place"),
(3) Run as Fast as You Can ("the next action is to run in place as fast as you can"),
(4) Jump as High as You Can ("the next action is to jump as high as you can"),
(5) Do Five Jumping Jacks ("the next action is to do five jumping jacks"), and
(6) Fly Like a Bird ("the next action is to fly like a bird").

The order of the actions to be performed by the actors was randomized. We selected these six actions to span a range of dynamic motions, and include common motions typically studied in similar perceptual experiments [Johansson 1973; Dittrich et al. 1996; Cutting 1978].

After the movement data had been recorded, we created point light displays of each actor performing each action. The point light displays consisted of round white markers at the captured joints of the actor, rendered on a black background, rendered out as a video over the duration of the action (the countdown from the T-pose was not included). Since the data had been captured at 30 frames per second (fps), the point light display was played back at 30 fps in the stimuli videos. All videos were rendered from the frontal view.

The Kinect records the three dimensional (3D) joint positions of twenty joints in the human body, $[x_i^j, y_i^j, z_i^j]$ where $j = 1, 2, \cdots, 20$ denotes the joint indices, and i indexes the frames of the recording. We represent the vector of joint positions for the frame i as $\mathbf{p}_i = [x_i^1, y_i^1, z_i^1, x_i^2, y_i^2, z_i^2, \cdots, x_i^{20}, y_i^{20}, z_i^{20}]^T$, where the $x - z$ plane is the ground plane, and the y-axis marks the height of the person.

To prevent obvious size differences between child and adult actors from cuing participants in the perceptual experiment, we scaled all stimuli to be the same physical height. This scaling was achieved by computing the mean value over all N frames and all twenty joints:

$$\mu_y = \frac{\sum_{i=1}^{N} \sum_{j=1}^{20} y_i^j}{20 * N}$$

Then, the joint positions were centered relative to this mean by subtracting μ_y from each y_i^j. The y positions of the head and foot were used to estimate the height of the actor, and the rendered point

Actions	Mean Time (seconds)			Mean No. Repetitions		
	Children [SD, *N* = 4]	Adults [SD, *N* = 4]	Overall [SD, *N* = 8]	Children [SD, *N* = 4]	Adults [SD, *N* = 4]	Overall [SD, *N* = 8]
Fly Like a Bird	4.7 [1.2]	5.6 [1.3]	5.1 [1.3]	5.75 [1.0]	5.50 [0.6]	5.63 [0.7]
Jump High	2.8 [0.8]	2.4 [0.4]	2.6 [0.7]	2.25 [1.5]	1.00 [0.0]	1.63 [1.2]
5 Jumping Jacks	5.0 [0.3]	5.8 [0.6]	5.4 [0.6]	5.00 [0.0]	5.25 [0.5]	5.13 [0.4]
Run Fast	3.5 [1.0]	4.1 [0.9]	3.8 [1.0]	9.50 [1.0]	10.25 [1.3]	9.88 [1.1]
Walk in Place	6.6 [1.1]	7.4 [1.0]	7.0 [1.0]	10.0 [0.0]	10.25 [1.3]	10.1 [0.8]
Wave	4.1 [0.9]	5.0 [1.0]	4.6 [1.0]	5.25 [0.5]	5.25 [0.5]	5.25 [0.5]

Fig. 3: Average and standard deviations of the duration (in seconds) of and number of repetitions in the point light videos in our stimuli set, separated by children and adults, and overall.

light display was scaled to a canonical height by zooming the render camera towards or away from the joint markers. We added padding above and below to account for extra space needed in actions such as “Jump as High as You Can”.

The “count-down” and ending position were cropped so that only the action itself was presented in the point light display stimuli. The durations of the actions differed slightly across actors. Figure 3 shows the average and standard deviations of the duration in time and number of repetitions for each action, for children and adults and overall. In general, children’s motions are marginally shorter than adults’ actions with respect to time duration, but this is not due to having done fewer repetitions. Instead, it seems that the children’s motions are more rapid than the adult motions in our stimuli set.

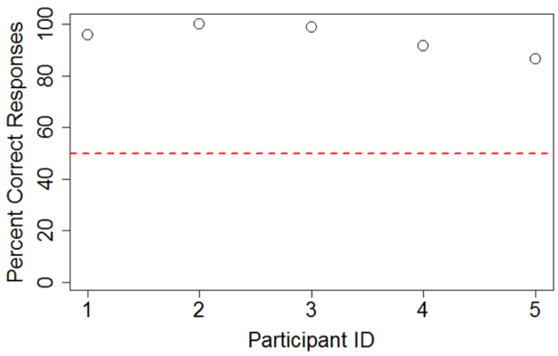
4. PRE-EXPERIMENT: ARE BOTH CHILD AND ADULT MOTIONS PERCEIVED AS DIFFERENT FROM RANDOM MOTION?

We conducted a pre-experiment study to verify that naive viewers perceive both child motion and adult motion as different from random motion. The goal of this pre-experiment is to check that the stimuli generated by us are not processed as random, or non-biological, motion. Verifying this would suggest that the participants in the main experiment (described next) are making their judgement of adult versus child based on the adult-like or child-like characteristics of the motion, respectively.

For this study, we recruited six adult participants who volunteered their time under a protocol approved by our Institutional Review Board (IRB). Data from one participant was incomplete (missing responses for 31 of 96 videos) and so was discarded. We had complete data for five participants (four female, age range 19 to 26 years, mean = 24.2, standard deviation = 3.0). All were students at our university, and all had normal or corrected-to-normal vision. The participants completed the study at their own leisure via an online survey mechanism.

The stimuli videos were presented one at a time, and participants were asked to make a forced-choice response as to whether the motion belonged to “a person” or “not a person.” All participants saw all stimuli videos (4 children + 4 adults = 8 actors * 6 actions = 48 stimuli videos). They also saw a corresponding equal number of stimuli with “random” motion. To create these random stimuli, we scrambled each of the 48 actor/action pairs as described next.

The mean position for each joint j over the duration of the action for N frames was computed for the x coordinate:



(a)

Person / Not-a-Person	Actors	Percent Correct Identifications (N = 60)
Adult (95%)	734	98%
	921	97%
	934	92%
	970	95%
Child (94%)	290	95%
	337	95%
	644	88%
	723	98%

(b)

Fig. 4: (a) Percentage correct responses for all participants that did the "Person or Not a Person" task. (b) All actors were almost equally well recognized.

$$\mu_{x^j} = \frac{\sum_{i=1}^{N} x_i^j}{N}$$

Similarly, the same computation was done for the y and z coordinates. Each joint marker was centered by subtracting the mean, and then shifted out by a random distance (r_x^j, r_y^j, r_z^j) between -0.5m and 0.5m in each dimension. For the x coordinate, for example, this would be $\tilde{x}_i^j = x_i^j - \mu_{x^j} + r_x^j$, and similarly, for the y and z coordinates. Thus, each point light marker would exhibit the same dynamics as the original human motion for that joint, but at a random location on the screen.

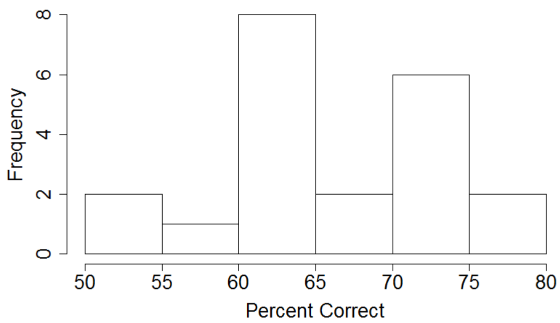
In total, participants in the pre-experiment saw 96 stimuli videos (48 person videos and 48 scrambled videos) in randomized order. They were asked to watch the video and then respond as to whether the motion in the video belonged to a "person" or "not a person". On average, it took participants about 53 minutes to complete the entire set, including some demographic questions at the end.

In addition to asking if the motion belonged to a person or not a person, we also asked participants if they could identify the action being performed. This question allowed us to ensure that the participants actually played the video since the study was administered online. All but one participant responded with reasonable labels indicating they viewed the videos (that participant's data was discarded for missing responses, as mentioned above).

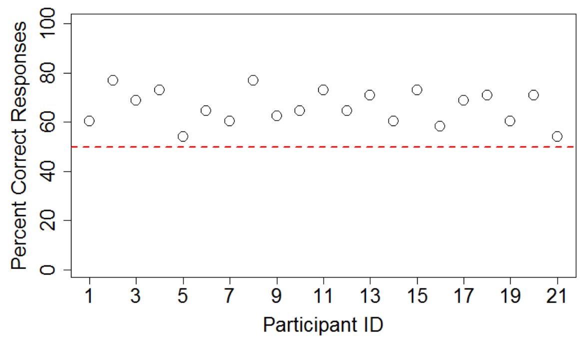
The responses to the two-alternative forced choice question were converted into an accuracy score for each participant by computing the percentage of correct responses per participant over all 96 videos. Figure 4a shows that all participants performed well above chance, with over 90% accuracy on this task. All actors were almost equally well recognized as exhibiting human motion (Figure 4b). This pre-experiment was a check on whether the recorded movement data was sufficiently different from random point light markers for both child actors and adult actors, and, due to the high accuracy rates, we considered the point light videos to be well suited for the main experiment.

5. MAIN EXPERIMENT: ARE NAIVE VIEWERS ABLE TO CORRECTLY IDENTIFY A MOTION AS BELONGING TO A CHILD VERSUS AN ADULT?

The main experiment we conducted focused on the question of whether or not naive viewers are able to perceive that child motion is different from adult motion when these motions are rendered as point



(a)



(b)

Fig. 5: (a) Percentage correct responses for all participants that did the "Child or Adult" task. (b) For each participant, the percentage of correct responses are shown as circle markers. The dotted line shows chance performance level. Three participants at or below chance level have been removed from further analysis.

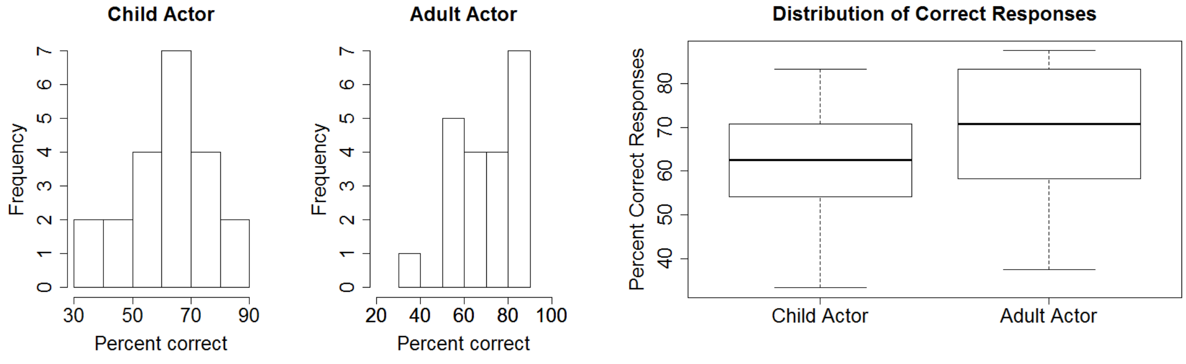
light displays. Our goal is to test if the different dynamics of child motion are perceivable when appearance and scale are not available as distinguishing cues. The findings from the pre-experiment indicate that child motion *is* perceived to be human motion. Therefore, any differences in responses in the main experiment are a result of viewers basing their judgements on the child-like and adult-like characteristics of the motion itself, rather than judgements of human or adult-like motion versus random or unfamiliar motion.

5.1 Study design

We recruited 34 adult participants who participated for extra credit in class under a protocol approved by our Institutional Review Board (IRB). Data from ten participants were incomplete (e.g., they did not answer one or more prompts during the survey or did not turn in their informed consent) and so were discarded. We had complete data for 24 participants (three female, age range 20 to 37 years, mean = 23.96, standard deviation = 4.4). All participants were students at our university, and had either normal or corrected-to-normal vision.

For each stimulus, each participant was presented with a video comprising an actor performing an action rendered as white point light markers on a black background. All videos were rendered at the same resolution, 854 $\times$ 480, at 30 fps. Because participants completed the study on their own computers via an online survey mechanism, the actual display size could not be controlled across participants. Participants were asked to play the video and answer the two-alternative forced choice question "Does this motion belong to a Child or an Adult?" Their response was recorded via a radio button that only allowed one option to be selected. Additionally, participants were asked to enter a freeform text answer to the question "What is the action being performed?" We used the responses to this question to ascertain that they had in fact played the videos, and identified the actions correctly. Seven participants had missing data for this question and were therefore discarded as mentioned.

Each participant saw all 48 stimuli videos (4 children + 4 adults = 8 actors * 6 actions = 48 stimuli). The order of stimuli videos was randomized for each participant. The median time to complete this survey was about 19 minutes; the average was 1 hour 9 minutes due to one outlier who did not complete the survey in one session.



(a) (b)

Fig. 6: (a) Distributions of percent correct accuracy for only the child videos, and only the adult videos, respectively. (b) Boxplots showing median and interquartile ranges of the accuracy for both groups of videos. Mean accuracy is significantly above chance for child videos and adult videos. The two group means are not significantly different.

5.2 Data analysis

Each participant provided an answer to the "Child or Adult" question for every stimulus video. The 48 responses received from a participant were converted into an accuracy value for that participant by computing the percentage of correct responses over all 48 videos. Figure 5a shows the distribution of accuracy across participants on this task. Figure 5b shows the accuracy for each individual participant. All participants are above the chance line. Note that we have removed three participants who had accuracy $\leq 50\%$. We judged these three participants to be outliers based on the performance of the rest of the participants who had a much higher accuracy at identifying child and adult.

Figure 6 shows the distributions of accuracy for the child actor videos and the adult actor videos, separately (Shapiro-Wilk tests indicated the distributions for both adult and child videos were normal, $p = 0.10 > 0.05$ and $p = 0.54 > 0.05$, respectively). We ran one-tailed t-tests to evaluate if the mean accuracy is significantly above chance in each category. Considering only the *child* actor videos, the mean accuracy (mean = 62.1%, standard deviation = 13.1%) was found to be significantly above the 50% chance level ($t = 4.23$, $df = 20$, $p < 0.05$). Considering only the *adult* actor videos, the mean accuracy (mean = 70.0%, standard deviation = 13.7%) was also found to be significantly above the 50% chance level ($t = 6.72$, $df = 20$, $p < 0.05$). The group means were not found to be significantly different from each other (paired t-test, $t = 1.59$, $df = 20$, $p = 0.13 > 0.05$).

Figure 7 shows the mean accuracy (separated by action) over all participants in identifying whether the presented point light stimulus video was an action performed by a child or an adult. The actions "Jump High", "Run Fast", and "Walk in Place" were all attributed correctly to child motion or adult motion with a high degree of accuracy. These motions are highly dynamic, and have previously been well studied in the context of identifying actions, gender, and emotion in point light based human motion perception studies [Johansson 1973; Cutting 1978; Dittrich et al. 1996; Cutting and Kozlowski 1977].

The largest difference in accuracy is observed for "Do 5 Jumping Jacks". Looking closer at the data for this action, we found that among the four child actors, the motion performed by actor 290 (a 6-year-old male) was correctly identified as belonging to a child by 18 out of 21 participants, whereas fewer than 60% of participants correctly identified the other child actors as children (meaning those actors

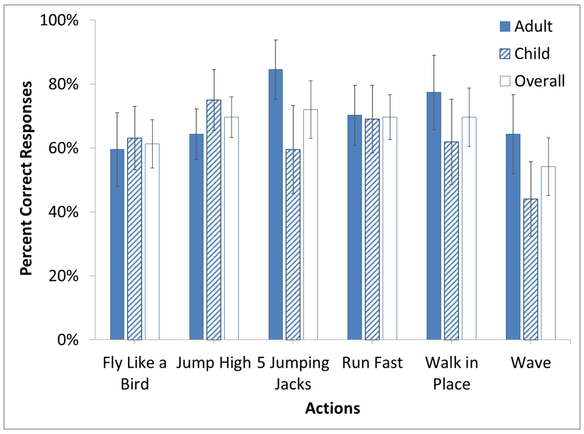


Fig. 7: Average accuracy over all participants in identifying whether an actor was a child or adult in our main experiment, separated by the action the actor was performing. Error bars are 95% confidence intervals.

were more likely to be identified as adults for this action). We hypothesize that this could be because viewers used coordination as a cue; the more uncoordinated the jumping jack action was, the more they identified it as belonging to a child. In the supplementary video, we show the four instances of “Do 5 Jumping Jacks” recorded from the child actors. Actor 290 (a 6-year-old male) is clearly less coordinated, while actor 337 (a 5-year-old female) is comparatively much more coordinated, and actors 644 and 723 (a 9-year-old female and 8-year-old male, respectively) are almost adult-like in their coordination. The videos of the adult actors doing jumping jacks are also shown as comparison.

The action “Jump High” has a fairly high recognition accuracy for both children and adults. This could be because the two younger children (actors 290 and 337) both did more than one jump, which might have been attributed by viewers to be a child-like thing to do. We found relatively low average recognition accuracies for the actions “Fly Like a Bird” and “Wave”. These actions are different from the rest because they primarily involve movement of the upper body while the feet are planted on the ground, and so are less dynamic. Possibly, these actions are also more open to interpretation. For example, for “Wave”, we found that there were instances amongst both adult and child actors in which the actor raised their hand above their head, or only to their shoulder, or only to their chest. This type of variability (in terms of how high or low the hands are held) occurred for “Fly Like a Bird” as well.

Figure 8 shows the distribution of accuracies on the “Child or Adult” question, considering the younger children and the older children separately (Shapiro-Wilk tests indicated the distributions for both younger and older children were normal, $p = 0.07 > 0.05$ and $p = 0.22 > 0.05$, respectively). Younger children were 5 to 6 years old, and older children were 8 to 9 years old. Each category has one male actor and one female actor. The mean accuracy in each case was found to be significantly above chance as follows. A one tailed t-test comparison to the 50% chance level for the *younger* child actor videos (mean = 64.7%, standard deviation = 16.9%) was significant: t = 3.99, df = 20, $p < 0.05$. A one tailed t-test comparison to the 50% chance level for the *older* child actor videos (mean = 59.5%, standard deviation = 15.4%) was also significant: t = 2.83, df = 20, $p < 0.05$. There was no significant difference in the group means (paired t-test, t = 1.25, df = 20, $p = 0.23 > 0.05$).

Figure 9 illustrates the accuracies of younger child actor videos versus older child actor videos, separated by actions. This analysis is preliminary because we have only two instances of younger child actors and two instances of older child actors. The trends in the graph indicate that motion performed by younger children is more readily identified as belonging to a child for the actions “Do 5 Jumping

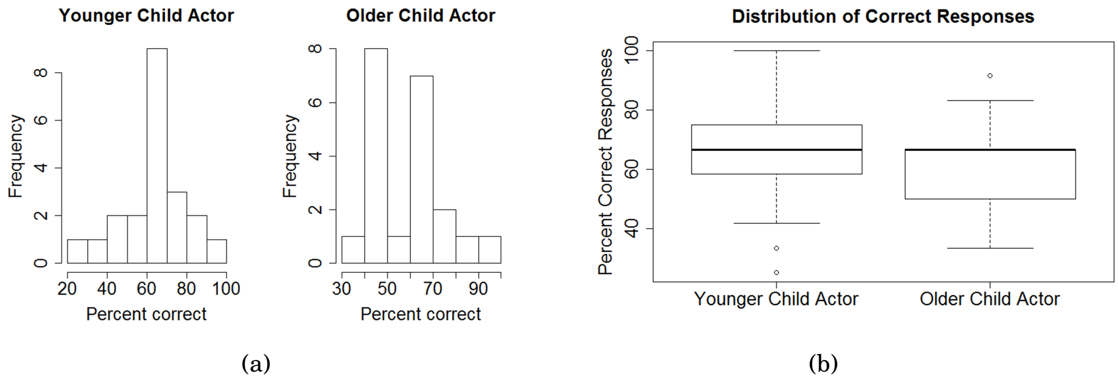


Fig. 8: (a) Distributions of percent correct accuracy in identifying whether an actor was a child or adult for only the younger child videos, and only the older child videos, respectively. Younger children were 5 to 6 years old, and older children were 8 to 9 years old. (b) Boxplots showing median and interquartile ranges of the accuracy for both groups of videos. Mean accuracy is significantly above chance for both age groups, but the two group means are not significantly different.

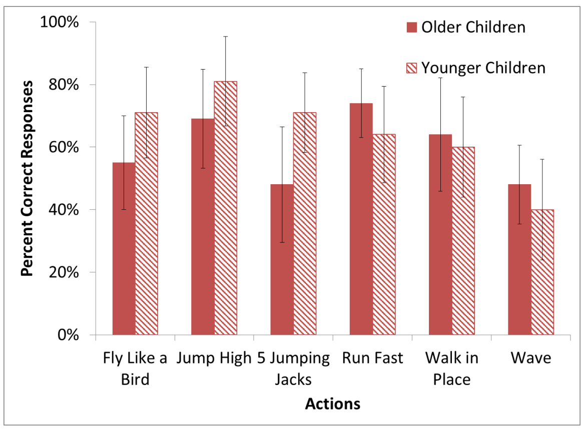


Fig. 9: Average accuracy for younger versus older children, separated by action. Error bars are 95% confidence intervals.

Jacks", "Jump High", and "Fly Like a Bird" as compared to "Run Fast", "Walk", and "Wave", which show less difference by actor age.

6. DISCUSSION

In this paper, we have presented an experiment to evaluate whether naive viewers can identify motion as belonging to a child or an adult from only point light markers placed at twenty joints on the actors' bodies. Our experimental setup involving the point light display paradigm, and a Kinect-based motion capture approach for recording the stimuli motion, was validated with a pre-experiment. The data from the main experiment indicates that people can identify if a motion was performed by a child or an adult at levels significantly above chance (50%); accuracy was 66% over the entire stimuli set of actions and actors. We note that previous work on perception of human motion via point light displays has reported success with similar levels of accuracy (e.g., [Barclay et al. 1978; Loula et al. 2005]).

Adult / Child	Actors	Percent Correct Identifications (*N* = 126)
Adult (70%)	**734**	72%
	921	73%
	934	69%
	970	66%
Child (62%)	**290**	68%
	337	62%
	644	67%
	723	52%

Fig. 10: Overall accuracy across all participants in identifying whether each actor is a child or an adult.

For dynamic actions involving the entire body (running, walking, jumping jacks), the recognition accuracy was about 70% when the analysis was over the entire stimuli set, i.e., considering child and adult actors together. Taken separately, the adults were on average more well recognized for walking than children, though these differences were not analyzed via formal statistical analysis. For actions that are expected to require coordination, we hypothesize that observers expect children to be less coordinated, and so the more coordinated jumping jacks were attributed to adult actors for both child and adult actor stimuli. Only one of the four children exhibited apparently less coordination in our stimuli set for this action, which could have led viewers to sometimes rate the child actors as adults in our results.

Children did the same number of repetitions as adults, but the durations were shorter on average. Figure 3 shows the average and standard deviations of the duration in time and number of repetitions for each action, for children and adults and overall. This figure indicates that for the children in our stimuli set, the motions are more rapid than the adults: the action durations are shorter, but the number of repetitions are approximately equal. This could be one of the factors that naive viewers use as a cue to identify whether the actor was a child or adult. While this is only a preliminary analysis into the velocity of an action, it would be consistent with the findings of Pollick et al. [2002] on the identification of gender from point light displays. Their work suggested that, among several factors that cause viewers to classify a point light display actor as male or female, the velocity of an action is particularly important.

A limitation of this study is that the viewers who participated were undergraduate and graduate students at our university, and may not have been as familiar with child motion as other segments of the population, such as children or the parents and teachers of small children. In the light of previous findings that viewers are able to more accurately recognize themselves and their friends relative to strangers [Cutting and Kozlowski 1977; Beardsworth and Buckner 1981; Loula et al. 2005; Wellerdiek et al. 2013], it is likely that children would be more sensitive to the motion of an avatar when it is actually generated from a motion-captured child actor. It would be interesting to revisit this study with young children as the naive viewers, and perhaps also their parents and teachers.

Another limitation of this work could be that the particular individuals who were our actors did not generate motions that were representative of general motion by children and adults. One effort to alleviate this concern was using four individuals in each category. We also looked at the overall accuracy among participants on identifying each actor as a child or adult (Figure 10). The percent correct accuracies indicate that the child actors 337 and 723 had lower accuracy and were perhaps

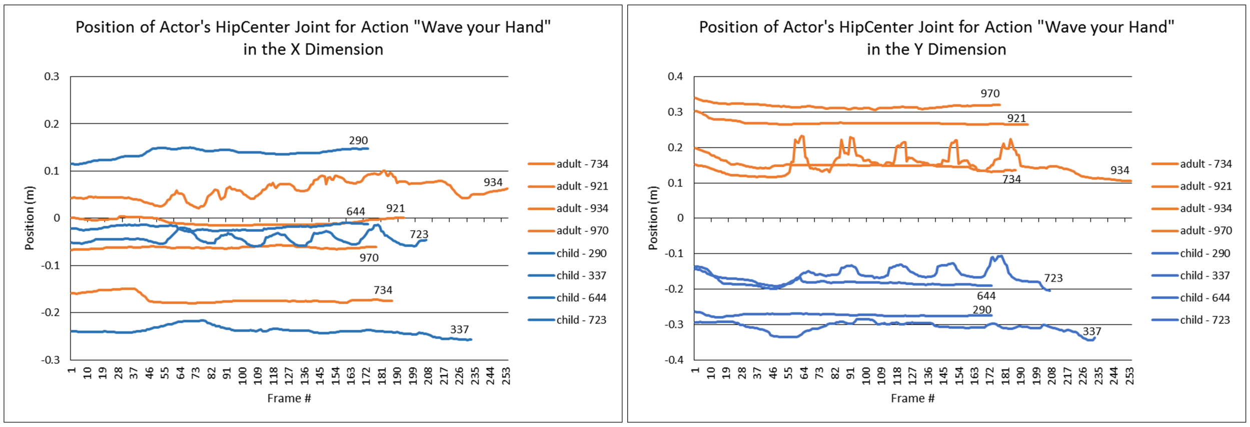


Fig. 11: A quality check on the child motion captured via Kinect. The variation in the position of a marker that should have ideally been stationary (the hip center joint, for the "Wave Your Hand" action) is similar for adults and children.

outliers in some way; the motion data would need to be examined in more detail to confirm this. We did not recruit trained actors so that we could elicit examples of the likely natural behaviors we might see in a system. Future work could consider using trained adult actors to, for example, perform the motion "as if" they were a child, to determine whether it would be sufficient to motion capture adults who are "pretending to be children." This is a strategy that is used in movies with variable success, e.g., the film *Polar Express* [Schaub 2005]. Previous research has shown that applying walking motion from a male actor onto a female character model (and vice versa) confuses viewers on a gender discrimination task [McDonnell et al. 2009]. An analogous perceptual study on the child-versus-adult axis might yield similar results. Also, we used only male adult actors, and given previous work which shows differences in motion by gender [Kozlowski and Cutting 1977], future work should use a balance of both genders for children and adults.

Although the heights of the point light displays were scaled to prevent absolute size from cuing participants, there could be other physical cues that were not normalized across our stimuli set. Examples of such cues include different body proportions between adults and young children such as relative head size, shoulder width, and leg to torso ratio [Huelke 1998]. Our experiment is a preliminary study, and it would be interesting future work to conduct follow up experiments that tease apart such factors, similar to the studies by Barclay and colleagues [Barclay et al. 1978].

Though the actions we selected represented motions commonly studied in the perception of human motion [Johansson 1973; Kozlowski and Cutting 1977; Dittrich et al. 1996; Hodgins et al. 1998], we investigated only six actions in this experiment. It would be a good idea to add more gestures and actions to future experiments to understand the space of potential variation among and between child and adult motion. Future work could also investigate numerical differences in motion characteristics such as joint accelerations, correlation between limbs, and variability in trajectory over repetitions to further our understanding of the underlying cues that viewers might use to differentiate between child and adult motion.

An important consideration for building a larger gesture and action set would be the quality of motion tracking obtained from markerless motion capture technologies, especially for children. We conducted a quality check on the child motion obtained using the Kinect in our study by considering the amount of "jitter" measured in a marker that should have been stationary. Figure 11 shows the x and y coordinates of the hip center marker for each of the child and adult actors for the action

"Wave your Hand". This marker should have been stationary because the lower body did not move in this action. There is a small amount of jitter in the measured positions, which can be quantified as the standard deviation of the position measurements. On average, the jitter is equivalent for adults ($\sigma_x = 0.009, \sigma_y = 0.014, \sigma_z = 0.005$), and for children ($\sigma_x = 0.009, \sigma_y = 0.012, \sigma_z = 0.008$). This is a sanity check rather than an exhaustive evaluation, but to this extent, the check suggests that the quality of child motion obtained via the Kinect is comparable to the quality of adult motion. Future work could consider direct comparison to other markerless motion capture systems.

7. CONCLUSION

In conclusion, this paper has reported the first explicit study of the perception of child motion (ages 5 to 9 years old), compared to analogous adult motion. The main contributions of this paper are (a) the application of the point light display paradigm towards studying the movement of young children and adults performing the same actions, and (b) the finding that naive viewers can identify a motion as belonging to a child actor or an adult actor in a two-alternative forced choice task at better-than-chance levels. This work has implications for creating more engaging and realistic avatars for games, online social media, and animated videos and movies.

Acknowledgements

The authors wish to thank University of Florida students Nathan de Krey and Zsolt Szabo, who both worked on early versions of the Microsoft Kinect skeleton tracking and data collection software, as well as several members of the HCC@UF community and our children helpers who served as pilot participants for both studies. The authors also thank the anonymous reviewers for suggestions that improved this paper.

REFERENCES

AnimationAddicts. 2013. The Challenge of Animating Underaged Adults. http://www.animation-addicts.com/2013/01/30/animating-kids/. (2013). Accessed: 2016-Feb-25.

Christine Assaiante. 1998. Development of Locomotor Balance Control in Healthy Children. *Neuroscience & Biobehavioral Reviews* 22, 4 (1998), 527–532.

Anthony P. Atkinson, Winand H. Dittrich, Andrew J. Gemmell, and Andrew W. Young. 2004. Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception* 33, 6 (2004), 717–746.

Catharine D. Barclay, James E. Cutting, and Lynn T. Kozlowski. 1978. Temporal and spatial factors in gait perception that influence gender recognition. *Perception & Psychophysics* 23, 2 (1978), 145–152.

T. Beardsworth and T. Buckner. 1981. The Ability to Recognize Oneself From a Video Recording of OneMovements Without Seeing OneBody. *Bulletin of the Psychonomic Society* 18, 1 (1981), 19–22.

Randolph Blake and Maggie Shiffrar. 2007. Perception of human motion. *Annual Review of Psychology* 58 (2007), 47–73.

Anna Brooks, Ben Schouten, Nikolaus F. Troje, Karl Verfaillie, Olaf Blanke, and Rick van der Zwan. 2008. Correlated changes in perceptions of the gender and orientation of ambiguous biological motion figures. *Current Biology* 18, 17 (2008), R728–R729.

Thierry Chaminade, Jessica Hodgins, and Mitsuo Kawato. 2007. Anthropomorphism influences perception of computer-animated characters' actions. *Social Cognitive and Affective Neuroscience* 2, 3 (2007), 206–216.

L.C. Chia, M.K. Licari, K.J. Guelfi, and S.L. Reid. 2013. A comparison of running kinematics and kinetics in children with and without developmental coordination disorder. *Gait & Posture* 38, 2 (2013), 264–269.

James E. Cutting. 1978. Generation of synthetic male and female walkers through manipulation of a biomechanical invariant. *Perception* 7, 4 (1978), 393–405.

James E. Cutting and Lynn T. Kozlowski. 1977. Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society* 9, 5 (1977), 353–356.

James W. Davis. 2001. Visual Categorization of Children and Adult Walking Styles. In *Proceedings of the Third International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA '01)*. 295–300.

Scott L. Delp, Frank C. Anderson, Allison S. Arnold, Peter Loan, Ayman Habib, Chand T. John, Eran Guendelman, and Darryl G. Thelen. 2007. OpenSim: Open-Source Software to Create and Analyze Dynamic Simulations of Movement. *IEEE Transactions on Biomedical Engineering* 54, 11 (Nov 2007), 1940–1950.

Winand H. Dittrich, Tom Troscianko, Stephen E. G. Lea, and Dawn Morgan. 1996. Perception of Emotion from Dynamic Point-Light Displays Represented in Dance. *Perception* 25, 6 (1996), 727–738.

Robert Fox and Cynthia McDaniel. 1982. The perception of biological motion by human infants. *Science* 218, 4571 (1982), 486–487.

Roberta Michnick Golinkoff, He Len Chung, Kathy Hirsh-Pasek, Jing Liu, Bennett I. Bertenthal, Rebecca Brand, Mandy J. Maguire, and Elizabeth Hennon. 2002. Young children can extend motion verbs to point-light displays. *Developmental Psychology* 38, 4 (2002), 604–614.

Jason Harrison, Ronald A. Rensink, and Michiel van de Panne. 2004. Obscuring Length Changes During Animated Motion. *ACM Transactions on Graphics* 23, 3 (Aug. 2004), 569–573.

Jessica K. Hodgins, James F. O. Brien, and Jack Tumblin. 1998. Perception of human motion with different geometric models. *IEEE Transactions on Visualization and Computer Graphics* 4, 4 (1998), 307–316.

Donald F. Huelke. 1998. An overview of anatomical considerations of infants and children in the adult world of automobile safety design. In *Annual Proceedings of the Association for the Advancement of Automotive Medicine*, Vol. 42. 93–113.

Yuri P. Ivanenko, Nadia Dominici, Germana Cappellini, Ambrogio Di Paolo, Carlo Giannini, and Richard E. Poppele. 2013. Changes in the spinal segmental motor output for stepping during development from infant to adult. *The Journal of Neuroscience* 33, 7 (Feb 2013), 3025–3036.

Gunnar Johansson. 1973. Visual perception of biological motion and a model for its analysis. *Attention, Perception, & Psychophysics* 14, 2 (1973), 201–211.

Ami Klin, David J. Lin, Philip Gorrindo, Gordon Ramsay, and Warren Jones. 2009. Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature* 459, 7244 (May 2009), 257–261.

Lynn T. Kozlowski and James E. Cutting. 1977. Recognizing the sex of a walker from a dynamic point-light display. *Perception & Psychophysics* 21, 6 (1977), 575–580.

Fani Loula, Sapna Prasad, Kent Harber, and Maggie Shiffrar. 2005. Recognizing people from their movement. *Journal of Experimental Psychology: Human Perception and Performance* 31, 1 (2005), 210–220.

Rachel McDonnell, Sophie Jörg, Jessica K Hodgins, Fiona Newell, and Carol O'sullivan. 2009. Evaluating the effect of motion and body shape on the perceived sex of virtual characters. *ACM Transactions on Applied Perception (TAP)* 5, 4 (2009), 20:1–20:14.

Rachel McDonnell, Sophie Jörg, Joanna McHugh, Fiona Newell, and Carol O'Sullivan. 2008. Evaluating the emotional content of human motions on real and virtual characters. In *ACM Symposium on Applied Perception in Graphics and Visualization (APGV)*. 67–74.

Philip R. Nader, Robert H. Bradley, Renate M. Houts, Susan L. McRitchie, and Marion O'Brien. 2008. Moderate-to-vigorous physical activity from ages 9 to 15 years. *Journal of the American Medical Association* 300, 3 (2008), 295–305.

Peter Neri, M. Concetta Morrone, and David C. Burr. 1998. Seeing biological motion. *Nature* 395, 6705 (1998), 894–896.

Marina Pavlova, Ingeborg Krgeloh-Mann, Alexander Sokolov, and Niels Birbaumer. 2001. Recognition of point-light biological motion displays by young children. *Perception* 30, 8 (2001), 925–933.

Frank E. Pollick, Vaia Lestou, Jungwon Ryu, and Sung-Bae Cho. 2002. Estimating the efficiency of recognizing gender and affect from biological motion. *Vision Research* 42, 20 (2002), 2345–2355.

Karl S. Rosengren, Frederik J.A. Deconinck, Louis A. DiBerardino III., John D. Polk, Jesse Spencer-Smith, Dirk De Clercq, and Matthieu Lenoir. 2009. Differences in gait complexity and variability between children with and without Developmental Coordination Disorder. *Gait & Posture* 29, 2 (2009), 225–229.

Ryusuke Sakuma, Atsuko Gunji, Takaaki Goto, Yosuke Kita, Toshihide Koike, Makiko Kaga, and Masumi Inagaki. 2012. Effectiveness of social skills training for children with developmental disorders: behavioral analysis using a two-dimensional motion capture system. *No to hattatsu. Brain and Development* 44, 4 (July 2012), 320–326.

Marie Sandlund, Kei Hoshi, Eva L. Waterworth, and Charlotte K. Hager. 2009. A Conceptual Framework for Design of Interactive Computer Play in Rehabilitation of Children with Sensorimotor Disorders. *Physical Therapy Reviews* 14, 5 (2009), 348–354.

David Schaub. 2005. 'The Polar Express' Diary: Part 3 – The MoCap/Anim Process. http://www.awn.com/vfxworld/polar-express-diary-part-3-mocapanim-process. (2005). Accessed: 2016-Apr-29.

Waltraud Stadler, Anne Springer, Jim Parkinson, and Wolfgang Prinz. 2012. Movement kinematics affect action prediction: comparing human to non-human point-light actions. *Psychological Research* 76, 4 (2012), 395–406.

Esther Thelen. 1995. Motor development: A new synthesis. *American Psychologist* 50, 2 (1995), 79–95.

Anna C. Wellerdiek, Markus Leyrer, Ekaterina Volkova, Dong-Seon Chang, and Betty Mohler. 2013. Recognizing Your Own Motions on Virtual Avatars: Is It Me or Not?. In *Proceedings of the ACM Symposium on Applied Perception (SAP)*. 138–138.