TAILORING MOTION RECOGNITION SYSTEMS TO CHILDREN'S MOTIONS

By

AISHAT ALOBA

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2021

I dedicate this dissertation to my loving parents, Engr. Rahman Aloba and Engr. (Mrs.) Musiliat Aloba, my siblings, and the Akinyodes for all their continued support, prayers, and words of encouragement.

ACKNOWLEDGEMENTS

I would like to specially thank my advisor, Dr. Lisa Anthony for her continued guidance and support, words of advice, and encouragement throughout my PhD. I especially appreciate the countless hours providing feedback on my research ideas and paper drafts, which helped to improve my research and writing skills, for sharing opportunities and resources to aid my academic and professional development, and the constant reminders to have a work-life balance. I would also like to thank all my committee members for agreeing to serve on my committee and for their continued support. To my internal committee members, Dr. Jaime Ruiz and Dr. Eakta Jain, I thank you for spending hours listening to my research ideas and providing valuable feedback. To my external member, Dr. Angelou Barmpoutis, I thank you for sharing valuable grant opportunities.

I would also like to thank the members of the INIT and Ruiz labs for always being a resource that I could rely on to discuss research ideas, proofread paper drafts, get valuable feedback, and share research and networking opportunities. In addition, I would also like to thank members of the Human-Centered Computing (HCC) community for providing an environment that fostered learning and collaboration. Finally, I would like to express my heart-felt gratitude to my family and friends for being a great support system, without which I would have been unable to complete this milestone.

TABLE OF CONTENTS

LIST OF TABLES

Tables page

LIST OF FIGURES

Figures page

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

TAILORING MOTION RECOGNITION SYSTEMS TO CHILDREN'S MOTIONS

By

Aishat Aloba

August 2021

Chair: Lisa Anthony
Major: Human-Centered Computing

Motion-based applications allow users to interact with a system or interface using motions. These applications often require accurate recognition of motions to ensure meaningful interactive experiences and are becoming increasingly popular among children. However, motion recognition systems are usually trained on adults' motions even though children move differently from adults. Our findings showed that naïve viewers could perceive the difference between children's and adults' motions at levels significantly above chance when the motion was abstracted from appearance cues (e.g., height). Our findings further showed that skeleton-based motion recognition systems, which accept as input the positions of joints tracked by a motion sensor in 3D space over time, perform poorly on children's motions compared to adults' motions. Therefore, motion recognition systems should be tailored to children's motion qualities to enable accurate recognition.

To characterize children's natural motion qualities, we focused on understanding how children perform motions. We designed a method to analyze the key body parts that users move during a motion and found that children perform motions more inconsistently compared to adults. To understand why children's motions are more inconsistent, we quantified the differences between children's and adults' motions. We initially analyzed children's and adults' walking and running motions using gait features and found that children move in less time and with higher energy. Then, we generalized this analysis to a broader set of motions by defining 24 articulation features (11 of which we newly identified), that quantitatively describe users' motions

performance. We analyzed these features on a subset of children's and adults' motions from the Kinder-Gator dataset to reveal new insights about how children perform motions (i.e., their motion qualities).

Based on our findings, we propose guidelines for tailoring motion recognition systems to children's motion qualities to enable accurate recognition of children's motions and guidelines for designing motion sets and motion applications for children. The implementation of these guidelines to tailor actual recognition systems to children's motions and evaluate their performance is out of the scope of this dissertation work. Finally, we also provide ideas for continued research in improving children's interactive experiences in motion-based applications.

CHAPTER 1
INTRODUCTION

The prevalence of low-cost tracking sensors that can accurately track users' movements, such as the Microsoft Kinect [89] and most recently, the Azure Kinect DK [90], has increased the popularity of applications that support whole-body gesture interaction (i.e., motion-based applications). For example, these sensors have enabled the development of exertion games that translate physical movements into game commands to induce physical exertion, such as Kinect Sports [109] and Just Dance 3 [131] and assistive robots that aid humans with tasks by relying on motions to facilitate human-robot collaboration, such as the Baxter robot [83]. To support whole-body gesture interaction, motion-based applications usually include motion recognition algorithms that can accurately recognize the specific sets of motions that the applications support (i.e., motion sets). Accurate recognition of motion sets plays an important role in users' interactive experiences. Prior work has found that the precision of motion recognition algorithms is positively associated with higher levels of immersion in exergames [98].

Motion-based applications are becoming increasingly popular among children as researchers and practitioners have started using these applications to target children's needs. For example, exergames, such as iFitQuest [81] and Vortex Mountain [141], were specifically designed to increase the time children spend engaged in physical activity. Researchers in the emerging field of child-robot interaction have also designed robots that can facilitate social interaction with children [12]. For example, Belpaeme et al. [12] designed a companion robot that interacts with children using verbal (speech) and non-verbal (body movement) gestures to help improve children's experiences during a hospital stay and learn more about their health conditions. However, motion recognition systems are usually trained on adults' motions [62, 105, 110, 153]. Children will likely move differently from adults due to differences in their body proportions and stages of neuromuscular development [59]. Therefore, motion recognition systems trained on adults' motions will likely perform poorly on children's motions. The goal of this dissertation work is to answer the following research questions:

a. What are the differences between children's and adults' motions?

b. What inferences can we make from children's and adults' motion differences to help tailor motion recognition systems to children's motions?

More specifically, our thesis statement is: "investigating the differences between children's and adults' motion will inform an understanding of children's natural motion qualities and allow the generation of guidelines for tailoring motion recognition systems to children's motion qualities for accurate recognition."

In this dissertation work, we began by investigating the perceptual differences between children's and adults' motions to show that children do indeed move differently from adults. Due to the obvious differences between children and adults that will likely affect their motions (e.g., body proportions), we believed that these differences will impact how they perform motions. Therefore, naïve viewers should be able to perceive the difference between children's and adults' motions. We conducted a perception study to investigate whether naive viewers could perceive the difference between children's and adults' motions, when the motion was abstracted from all appearance cues (e.g., face, height, and build) [64]. We found that naïve viewers could perceive this difference at levels significantly above chance, thus establishing that children move differently from adults.

To show that these differences will result in poor recognition performance of children's motions, we evaluated the performance of three motion recognition algorithms that have been shown to achieve good recognition performance ($>80\%$) on adults' motions [22, 5]. For the evaluation, we used a subset of children's and adults' motions from the Kinder-Gator dataset [3]. The Kinder-Gator dataset, which I helped collect, is a dataset of 10 children and 10 adults performing motions forward-facing the Kinect v1. We tested two template-based recognizers, Dynamic Time Warping (DTW) and a motion recognizer that we adapted from template-based stroke gesture recognizers [5], and a machine-learning recognizer using Support Vector Machines (SVMs) [22]. All three algorithms compare the motion to be recognized (candidate motion) to a set of motions selected for training (template motions) and select the motion from the training set that is the most similar to the candidate motion. Our findings showed that all algorithms had

higher recognition accuracies for adults' motions (DTW: 73.6%, template-based adaptation: 81.4%, SVM: 81.4%) compared to children's motions (DTW: 60.7%, template-based adaptation: 55.0%, SVM:70.0%). These findings showed that motion recognition algorithms should be tailored to children's natural motion qualities to enable accurate recognition of children's motions. However, the motion qualities that distinguish children's motions from adults'motions are not known, which makes tailoring motion recognition systems to children's motion qualities difficult.

After seeing the results from our perception study, we know that there are perceivable cues that differentiate child motion from adult motion. However, what exactly is being perceived has not been identified: what are the differences between child and adult motion? If these differences could be identified, then we can characterize children's natural motion qualities. In this dissertation work, we focused on understanding the nuances in how users articulate motions, which we propose can be achieved by a) identifying the joints that are critical to performing motions to investigate variations in how children and adults move their body parts when performing said motions and b) identifying features that can quantify how the motion is produced to aid in the analysis of child and adult motion. We designed a method, which we call "filterJoint", that selects the key joints that users are actively moving during the performance of a motion [5]. We operated on 'joints' because our method works on motion that was tracked using skeleton-based motion sensor technology, such as the Microsoft Kinect sensor [89] or the Vicon motion capture sensor (Oxford Metric, Oxford UK). These sensors track the motion of users' joints in space over time. Our focus on actively moving joints stemmed from the idea that both tracking noise from the sensor and unintentional movements from the user can affect how each joint is moved (e.g., joints that are not supposed to move will appear to move). Including such joints during analysis of motions could lead to incorrect inferences about how users perform motions. We compared the joints selected by our filterJoint method on a subset of children's and adults' motions from the Kinder-Gator dataset and found that children are more inconsistent in how they perform motions compared to adults [5].

Although our findings established that children are inconsistent when performing motions compared to adults, we are yet to understand why these inconsistencies occur because the differences between children and adults are yet to be quantified. If we can quantify these differences, then we can gain new insights into how children move (i.e., children's natural motion qualities). For example, we will be able to understand how and why children move inconsistently compared to adults. To help quantify these differences, we begin by identifying features that can aid with the analysis of walking and running motions. We initially concentrated on these motions because we found from our perception study that naive viewers could differentiate these motions with even higher accuracy compared to the other motions we selected for the study. We used temporal (time) and spatial (distance) features from the gait literature that has been used to characterize walking and running motions. We analyzed walking and running motions from the Kinder-Gator dataset [3] and found that children generally complete walking and running motions in less time and with higher energy compared to adults [4].

However, gait features are not generalizable to a broader set of motions (e.g., "Jump") due to their reliance on the periodicity of the motion. Hence, we identified a set of "articulation features" to reveal new insights into the ways children perform motions. An articulation features is any measure that quantifies a specific property associated with how a user performs a motion, such as length, shape, and time. We used features from prior work [132] that describe whole-body motions on a global level (i.e., dependent on the postures that make up the motion). However, because these features focus on the whole body, they will not be useful for characterizing properties of individual joints that are critical to performing motions. For example, when performing a "Jump" motion, we expect users to lift both feet off the ground. Depending on how users control the joints in their feet, there could be variations in how users lift their feet off the ground, which could result in differences in how these users perform the "Jump" motion. In this example, global level features will be useful in characterizing whether users are performing the "Jump" motion as opposed to a different motion (e.g., "Bend your knee" motion), but these features will not be useful in characterizing whether users moved individual joints that are

responsible for performing the jump motion, such as limb movements [92] differently. To characterize how users move individual joints, we identified a set of joint-level features (i.e., dependent on the joints tracked by the motion sensor) and analyzed these features on a subset of children's and adults' motions from the Kinder-Gator dataset[3]. Our results showed that children move differently from adults in ways that can be quantified with specific posture- and joint-motion-based articulation features. For example, we found that children require space (proportionally) and show more variations in how they move body parts to perform motions, as compared to adults. To identify the optimal set of features for differentiating children's motions from adults' motions, we used a feature selection approach that relies on the classification accuracy of a machine learning recognizer. In this approach, we trained and tested the classifier on the same subset of children's and adults' motions, in which each motion is represented in the form (x,y) where x represents a set of articulation features and y represents the true label in which the motion belongs (i.e., child, adult). We repeat the training and testing process and select the subset of features that achieves the highest accuracy as the optimal set of features (more details in Chapter 6). The optimal set of features achieved an accuracy of 81%, further indicating that our articulation features are effective in the automatic classification of motions by age group (i.e., child vs. adult).

To provide further support for our quantitative results, we qualitatively analyzed children's motions to understand in what ways their motions are inconsistent from adults. During the analysis, we also explored the differences between younger and older children's motions since prior work in motor development literature [23] has shown that age plays an important role in the development of children's motor skills, which will impact how children perform motions. We used an inductive thematic approach [87] for the analysis, from which we revealed new findings about children's motion performance, further discussed in the dissertation.

Based on findings from our quantitative and qualitative work, which we synthesized to reveal new insights about how children perform natural motions, we introduce a set of new design guidelines for tailoring existing motion recognition systems to children's motion qualities for

accurate recognition of children's motions.The guidelines proposed from our findings have the potential to improve the performance of motion recognition systems on children's motions since prior work has found several correlations between global-level features and the performance of motion recognition systems on adults' motions [132] and 2D stroke gesture research has also shown a correlation between articulation features and children's stroke gesture recognition performance [120]. The implementation of these guidelines to tailor actual recognition systems to children's motions and evaluation of their performance is out of the scope of this dissertation work. We also propose guidelines for designing motion sets and motion applications for children. We hope that researchers and designers will adopt these guidelines to improve children's interactive experiences with motion applications. Finally, we also provide ideas for continued research in improving children's interactive experiences in motion-based applications and describe potential future research opportunities based on findings from our studies.

1.1 Overview of Work

This dissertation aims to characterize the differences in children's and adults' motions to propose guidelines that future research can use to tailor motion recognition systems to children's motion qualities. The contributions of this dissertation include:

- Introduction of the Kinder-Gator dataset, a publicly available dataset containing the motions of 10 children and 10 adults, to enable comparison of their motions [1]
- Established that naïve viewers can perceive the difference between child and adult motion at levels significantly above chance when the motion is abstracted from appearance cues [2]
- Comparison of the recognition rates for children's and adults motions when both motions are trained and tested independently on two skeleton-based motion recognition systems.
- Analysis of children's and adults' walking and running motions using time-based and distance-based features from existing literature on the analysis of gait to establish a set of features that quantify the differences between how children and adults walk/run.

[1] I acted as the first experimenter during the collection of half of the adult participants' motions during which I was responsible for informing the participant about the next motion to be performed and recording the motion as the motion is performed. In addition, I was the first author during the publication of the dataset [3], responsible for writing the paper and leading a team of students to organize the dataset for publication.

[2] I was responsible for abstracting the motion from all appearance cues (e.g., creating the point light display videos) and analyzing the data collected to extract the results. For the publication [64], I designed the box plots and scatter plots and wrote a section detailing how the motions were abstracted.

- Implementation of an automated approach that selects only the key joints that are actively moving joints during the performance of a motion and a presentation of how these joints can facilitate an understanding of children's and adults' motion articulations.
- Identification of a set of joint-level features (i.e., depends on the joints tracked by the Kinect) that characterize how children and adults move their joints when performing motions.
- Analysis of children's and adults' motions using global-level features (i.e., dependent on the posture) proposed by Vatavu [132] and our proposed joint-level features to quantify the similarities and differences in how children and adults perform motions.
- Identification of an optimal set of articulation features for accurate classification of motions by age group (i.e., child vs. adult)
- Identification of a set of qualitative themes explaining how children perform motions.
- Presentation of a set of design guidelines for future work to help design recognizers that can achieve higher recognition rates on children's motions.

1.2 Dissertation Outline

The rest of this dissertation is organized as follows. Chapter 2 provides an overview of the literature on children's motor development and human motion recognition with emphasis on skeleton-based human motion recognition. Chapter 3 presents the scope of this work, detailing topics, such as the dataset and the population. Chapter 4 presents studies to establish a need to understand children's motions. Chapter 5 presents initial studies investigating children's and adults' motion articulations. Chapter 6 focuses on articulation features and motion descriptors that can help to characterize children's natural motion qualities. Chapter 7 introduces design guidelines for tailoring motion recognition systems to children's motions based on the findings from previous chapters. Chapter 8 concludes the dissertation, detailing the contributions and future work.

CHAPTER 2
BACKGROUND

In this chapter, we present a detailed overview of human motion recognition with emphasis on the design and evaluation of skeleton-based recognition algorithms. Because the goal of our work is to tailor motion recognition systems to children's motions, which is closely related to prior work in stroke gesture research, this chapter will also present a review of prior work in stroke gesture recognition.

2.1 Human Motion

In motion recognition research, human motion has been defined at different levels of abstraction depending on what the researcher aims to understand from the motion. Bobick [16] reviewed approaches used in computer vision for analyzing motions and proposed three levels for understanding motions based on the nature and amount of knowledge required to understand the motion. The first level is "movements", which are atomic primitives, such as performing a forehand shot in a tennis match, that require no contextual knowledge to be recognized [132]. The second level is "activities", which are sequences of movements for which understanding the properties, such as variability of appearance [16], of the sequence is important. The last level is "action", which the author defines as large-scale events that typically include interactions with the environment and causal relationships [16]. Human motions have also been categorized conceptually based on the complexity and number of body parts involved into "gestures", "actions","activity", and "group activities" [29, 66]. "Gestures" are movements of a person's body part that convey meaning, for example, "Lifting a leg" [1]. "Actions" are a collection of gestures performed by a single person (e.g., "Walking"). "Activity" are a collection of actions performed by at least two people (e.g., "Two people fighting"). "Group activities" are a combination of gestures, actions, or interaction involving multiple groups of people (e.g., "Two groups of people marching")[29, 1].

Because of these different levels of abstractions (e.g., nature, complexity), throughout this dissertation work, we define human motions as whole-body movements that contain information. This definition follows the principle of Kurtenbach and Hulteen [71] who asserted that the difference between a gesture (i.e., motion) and generic movement is dependent on the information

it carries: “A gesture is a motion of the body that contains information. Waving goodbye is a gesture. Pressing a key on a keyboard is not a gesture because the motion of a finger on its way to hitting a key is neither observed nor significant. All that matters is which key was pressed.” This movement ranges from simple limb movements, such as “Wave” and “Raise a hand”, to movements involving multiple limbs of the human body, such as “Walk” and “Jumping jacks.”

2.2 Human Motion Recognition

Human motion recognition, also referred to as “whole-body gesture recognition” or “action recognition”, is a systematic approach to understand and analyze human motion [143]. Human motion recognition has many application areas that spread across a variety of domains, including games (e.g., exergames), human-robot interaction (e.g., assistive robots), and security (e.g., smart surveillance systems):

Exertion games. Exertion games, also known as exergames, combine videogames with physical activity [80, 141]. Physical activity usually involves the performance of certain motions, so exergames must support whole-body gesture interaction (i.e., interaction through motions). To support whole-body gesture interaction, exergames include motion recognition systems that can accurately classify whole-body gesture sets that the exergame supports.

Assistive robots. Assistive robots are designed to help relieve humans from heavy tasks in home and industry settings through human-robot interaction. To facilitate human-robot interaction between humans and robots, effective communication is important [77]. Since motions are effective communication tools in human-human interaction, assistive robots usually rely on motion recognition systems to communicate with humans [77].

Smart surveillance systems. Surveillance systems that monitor people and events through videos are used for maintaining security in the real-world. Smart surveillance systems are systems that can “preempt incidents through real-time alarms of suspicious behaviors” [52]. To achieve this, smart surveillance systems must be able to automatically analyze visual surveillance videos. Since motion is the most important cue for identifying dynamic content in videos [58], smart surveillance systems must include accurate motion recognition systems to analyze motions in

videos to identify suspicious behaviors. Human motion recognition has been studied actively for more than two decades [66], primarily in the field of computer vision and human-computer interaction (HCI). Research on human motion recognition has focused on developing accurate and effective frameworks for human motion recognition [68]. Recognition of human motion has been studied extensively using traditional computer vision approaches that rely on using pixels to recognize motions from images and videos (i.e., vision-based human motion recognition) [32, 111]. For example, Dreijer and Herbst subtracted each pixel in consecutive frames of a video and averaged over all consecutive frames to create a new image that represents the average changes in pose [32]. The researchers evaluated their approach on the KTH dataset [115], which includes videos of 25 participants performing six actions ("Walking", "Running", "Jogging", "Hand waving", "Boxing", and "Clapping") and found an accuracy of 86.0% using nearest neighbor classification and 87.8% using Support Vector Machines (SVMs). Rodriguez et al. [111] created a gesture recognizer that relies on a maximum average correlation height (MACH) filter to recognize gestures from videos. The researchers found that their recognizer achieved up to 88.7% accuracy on the KTH dataset. For accurate recognition, vision-based approaches require the identification of the human from the video. However, prior work has noted that this identification can be challenging, especially in the presence of factors such as background clutter, illumination changes, and occlusion [33, 93].

With the advent of motion capture (mocap) devices that track the movement of the body in space over time, researchers have shifted focus from vision-based human motion recognition to skeleton-based human motion recognition. Skeleton-based human motion recognition relies on the information provided by the mocap device to analyze and recognize human motion data [154]. Mocap devices provide accurate information about human body movement (e.g., posture information of the human body defined by the positioning of the body at a specific time instance). For example, the Microsoft Kinect tracks the movement of 20 joints in 3D space over time. For this dissertation work, we will focus on skeleton-based human motion recognition. In contrast to vision-based approaches that need to separate the human from the background in the video to

characterize the motion being performed, skeleton-based approaches eliminate this additional step as the mocap devices already provide motion details about the human body [154]. The rest of our review on human motion recognition will focus on skeleton-based human motion recognition.

2.3 Skeleton-Based Human Motion Recognition

Skeleton-based human motion recognition relies on properties of the skeleton information (e.g., posture information) provided by the mocap device to analyze and recognize human motion data [154]. In skeleton-based human motion recognition research, whole-body motions usually consist of a set of body poses or postures unfolding in time such that each pose includes a set of joints representing the skeletal structure of the human body. Human motion recognition algorithms usually involve a two-step process: a) motion representation and b) motion classification.

2.3.1 Motion Representation

The first step in recognizing human motions is to extract features that contain discriminative information (e.g., spatio-temporal information) from the motion data, thus converting from data space to feature space [125]. In skeleton-based human motion recognition, prior work has extracted features from the motion of body parts [65, 99], which is a segment that connects two joints, for example, the forearm is a body part that connects the wrist and elbow joints [21]. For example, Nirjon et al. [99] extracted 22 features from the motion of body parts. These features include the orientations of two upper body parts (upper arm and thigh) defined as cosine angles, relative positions of two lower body parts (lower arm and leg) defined as the angle between the lower body part and their corresponding upper part, and unit normals of the planes where the body parts are located. This representation provides a general understanding of the motion being performed. However, global representation ignores local structure information (e.g., movements of specific joints), which can make recognition of similar motions challenging [33]. To address this problem, researchers have extracted features from the local joints tracked by the motion sensor (e.g., the wrist joint). A common feature extracted from the joints is the raw positional data. For example, Martinez-Zarzuela et al. [86] used the raw motion trajectories of 15 joints to

represent the motion. The joint trajectory is a set of 3D points showing the positions of the joint throughout the motion. Cippitelli et al. [22] and Duarte et al. [34] also used the raw input data but focused on the skeleton frames provided by the Kinect instead of motion trajectories. Because joint positions are not scale-invariant (i.e., the position is dependent on the height of the user), researchers have used joint angles instead [93]. For example, Th Papadopoulos et al. [104] represented motions in terms of the spherical angles of the relative positions of joints in the upper limb, such as the left shoulder and right wrist, and joints in the lower limb, such as the left knee and right foot. Ofli et al. [102] represented motions in terms of the 21 joint angles (e.g., angle of the right shoulder and angle of the left knee).

Although research has primarily focused on either a global level or local level representation, some researchers [21, 50, 154] have also combined both representations to benefit from their advantages. For example, Chikhaoui et al. [21] extracted global features from skeleton data by computing the distances between body parts (e.g., forearm) as well as relative and absolute joint angles between body parts. The authors extracted local features by computing pairwise distances of 3D joint positions at each time frame. Hachaj et al. [50] represented human motion in terms of angles between body parts (global) and joint angles (local). The global features extracted are the angles between forearms, the angle between vector defined by the joint between shoulders and the joint between hips and thighs, and the angle between thighs while the local features extracted are the joint angles of the elbow, shoulder, and knee joints.

Because features from these representations are usually applied to each frame or posture of the motion with each frame consisting of multiple joints or body parts, the number of features representing the entire motion can be very large. A problem resulting from this large number of features is that it may not accurately capture discriminative information about motion types [86]. To address this problem, researchers usually employ feature selection and dimension reduction techniques as an additional step when extracting features to improve the discriminative capability of these features. For example, Chikhaoui et al. [21] and Martinez-Zarzuela et al. [86] applied Singular Value Decomposition [43] and Fast Fourier Transform [24] respectively to select the

most discriminative features, while Xia et al. [153] applied Linear Discriminator Analysis (LDA) to extract the dominant features after feature extraction.

However, the reliance on these complex techniques usually results in features that are difficult to reason about intuitively [17] because they are tailored to algorithmic understanding (i.e., machine-readable). Therefore, researchers and practitioners will likely find it difficult to use machine-readable features to understand how users actually perform motions even though such information is important for improving the performance of motion recognition systems: prior work has noted that the differences in how users perform similar postures of a motion can negatively impact recognition accuracy [93]. In contrast, features that are tailored to human understanding (i.e., human-readable) will be informative in understanding how users actually perform motions [132]. To the best of our knowledge, only Vatavu has sought to propose features that characterize human motion performance [132] from a human-readable perspective. Vatavu proposed spatial features (e.g., features related to the amount of space required to perform the motion), kinematic features (e.g., features related to the time it takes to produce the motion), and body posture appearance features (e.g., features related to the deviation of the body posture from the centroid posture) as features for understanding users' whole-body gesture performance. These features rely on the overall posture of the body and are therefore global-level features. To combine the advantage of global- and local-level representations, we will also identify local-level features that characterize human motion performance in this dissertation work. These features (mine and Vatavu's [132]) will be fully described in Chapter 6.

2.3.2 Motion Classification

This step involves using algorithms that rely on the features extracted from the motion representation step to identify the motion type to which the motion data belongs. Classification algorithms for human motion recognition can be divided into two main types: template matching algorithms and machine learning algorithms. For each motion classification type, we discuss common algorithms used for human motion recognition followed by examples of the algorithms' recognition performance on publicly available dataset. Table 2-1 provides examples of publicly

available datasets that are commonly used to evaluate motion recognition algorithms. Because the evaluation of motion recognition algorithms is dependent on their performance (e.g., accuracy) on action datasets. To aid the discussion of classification algorithms for human motion recognition, we define the following terms: a) class: the motion type to which a motion data belongs (e.g., "Raise your hand"), b) candidate motion: the motion whose class is to be determined, c) template motion/training motion: a motion selected for training the classifier and whose class is known, and d) accuracy: ratio of correctly matched candidate motions to the total number of recognitions attempted. Both candidate and template motions are defined by features extracted from the motion representation set.

2.3.2.1 Template matching algorithms

Template-based algorithms directly match the candidate motion to a set of pre-stored template motions using point-by-point correspondence. Based on some distance metric (typically Euclidean distance), template matching algorithms select the template motion that is most similar to the candidate motion. The most popular template-matching algorithm for classifying human motions is Dynamic Time Warping (DTW). DTW measures the similarity between time sequence of motion measurements or features extracted from the motion measurements (e.g., angle vectors of each body part in a skeleton) [19]. The goal of DTW is to find the optimal alignment between two sequences that may vary in time or speed [78], which usually involves speeding up or slowing down a sequence in time [19]. Given a candidate motion and a template motion represented as sequences, DTW constructs a 2D matrix whose size is the product of the number of frames in the motions being compared. Each cell in the matrix is the matching cost, which is computed using dynamic programming and a distance function. In the computer vision and Human-computer Interaction (HCI) fields, Euclidean distance is typically used as the distance function [19], but other distance functions, for example, Mahalanobis distance [88], have been used. DTW finds the best match for the candidate motion by finding the template motion out of the set of pre-stored template motions that minimizes the cumulative matching cost required to align to the candidate motion sequence [19].

Table 2-1. Examples of publicly available datasets commonly used to evaluate skeleton based human motion recognition algorithms. Motions in the datasets are tracked using mocap sensors. All the datasets presented focus on only adults' motions. In the next chapter, we will present the Kinder-Gator dataset, which I helped collect. The Kinder-Gator dataset [3] is a publicly available dataset of children's and adults' natural motions.

Dataset	Description	Example Motions
MSR-Action3D dataset [75]	A dataset of 20 actions from 12 adult subjects tracked using the Kinect	High arm wave, forward punch, high throw, draw x, jogging, tennis swing.
MSRC-12 Kinect Gesture dataset [37]	A dataset that contains 12 actions performed by 30 adult subjects forward-facing a Kinect sensor. The dataset comprises of video and skeletal data.	Iconic gestures (e.g., Crouch or hide, Shoot with a weapon, Kick to attack an enemy). Metaphoric gestures (e.g., start music, navigate to next menu, take a bow to end the session)
HDM05-Mocap dataset [96]	A dataset that contains 100 actions performed by 5 adult subjects using a Vicon sensor (Oxford Metrics, Oxford UK). The dataset includes videos and skeleton data	Walking, dance, punch, kick, throw, squats.
UTD-MHAD dataset [20]	A multimodal dataset containing 27 actions performed by 8 adult subjects (4 males). Each subject repeated an action 4 times. The dataset comprises RGB videos, depth videos, and skeleton data from Kinect as well as data from a wearable inertial sensor.	Sports actions (e.g., bowling, tennis serve). Hand gestures (e.g., draw x, draw circle) Daily activities (e.g., knock on door, sit to stand). Training exercises (e.g., lunge, squat).
Florence-3D dataset [116]	A dataset that contains 9 actions performed by 10 adult subjects. Each action was performed twice or three times in front of a stationary Kinect sensor.	Wave, drink from a bottle, answer phone, clap, tight lace, sit down, stand up, read watch, bow.
UTKinect dataset [153]	A dataset that contains 10 actions performed by 10 adult subjects. Each action was performed twice in front of a stationary Kinect sensor. The dataset comprises RGB frames, depth frames, and skeleton data.	Walk, sit down, stand up, pick up, carry, throw, push, pull, wave and clap hands.

DTW is efficient in comparing motions performed at different speeds [19] and has been shown to achieve high recognition performance on motion datasets both in the computer vision and HCI fields [62, 110]. Riofrío et al. [110] used DTW to recognize 6 upper limb motions ("Right hand sweep left", "Right hand sweep right", "Right hand up", "Left hand up", "Two hands up", and "Two hands closed") articulated by adults and tracked using the Kinect. Because this dataset is not publicly available, it is not included in Table 2-1. The positions of 10 upper limb joints were selected as features and the algorithm achieved recognition accuracy of 84.07%. Similarly, Huu et al. [62] used DTW to recognize actions performed by adults and tracked using the Kinect. However, instead of selecting the class of the template motion that minimizes the Euclidean distance, the authors used a voting algorithm. The voting algorithm accepts the Euclidean distances from comparing training motions to a candidate motion as input and returns the class label as output. The relative angles between body parts were selected as features. The algorithm achieved recognition accuracy of up to 83.82% and 93.91% on the MSRAction3D dataset (see Table 1) and a dataset collected by the authors (not publicly available), which consists of six motions ("Left hand low waving", "Right hand low waving", "Left hand high waving", "Right hand high waving", "Hand clapping", and "Greeting"), respectively.

However, because DTW relies on a 2D matrix for its implementation, it is computationally expensive, thus making it unsuitable for real-time recognition [78]. To address this, researchers have designed template-based algorithms that require lower computational costs. For example, Jacknife [127] is an algorithm based on DTW that uses a rejection threshold to reduce the number of template motions being compared to the candidate motion, thus decreasing the computational cost. Wobbrock et al. designed the $-family of recognizers, for example, $1 [147], $N [8] for 2D touchscreen stroke gestures. Kratz and Rohz [69] created the $3 recognizer which extended the $-family of recognizers to 3D and enabled recognition of in-air hand gestures. This family of algorithms resamples the number of frames in a template and candidate stroke gesture (or motion) to the same size, which enables pairwise comparisons between the frames of both motions. Since the algorithm performs a pairwise comparison, the computational cost is an order of magnitude

less than that of DTW. In Chapter 5, we will show how the $-family template-matching algorithms can be adapted to enable accurate recognition of whole-body motions.

2.3.2.2 Machine learning algorithms

Traditional machine learning algorithms rely on statistical models to capture the unique characteristics of each motion type [78]. A set of pre-labeled training motions are used to estimate the parameters of a statistical model. The statistical model is usually referred to as a classifier [78]. Given a candidate motion, the classifier returns a discrete value corresponding to the predicted class to which the motion belongs. Common traditional machine learning algorithms for human motion recognition are Support Vector Machines (SVMs) [22, 77], K-Nearest Neighbors (KNNs) [152], Decision Trees [40], and Hidden Markov Models (HMMs) [153].

Support vector machines (SVMs) use planes to separate the set of training motions into their respective classes. The difference between support vectors (training examples closest to a hyperplane [77]) and a line is computed to define the margin of separation. The plane from amongst the planes that maximize the margin is referred to as the optimal plane and is used to determine the decision boundary for classification. When the training motion can be separated using a plane, the motions are said to be linearly separable [26]. In contrast, when the training motions are not linearly separable, SVMs introduce kernel functions, which transforms low-dimensional training data to higher dimensional feature space (e.g., linear kernel [77]). Then, an optimal hyperplane that separates the training motion in higher dimensions is used to create the decision boundary for classification [78]. Hussein et al. [61] designed a recognition algorithm that uses a covariance matrix representation of the skeleton joint locations and SVM as the classification algorithm. The algorithm achieved recognition accuracies of up to 90.53%, 98.7%, and 95.41% on three publicly available datasets: MSR-Action3D dataset, MSRC-12 Kinect Gesture dataset, and HDM05 mocap dataset (see Table 2-1), respectively.

K nearest neighbors (KNNs) approach classifies the candidate motion based on the closest K training motions (neighbors) [77, 107, 152]. The number K is a discrete value denoting the number of neighbors used to predict the class of the candidate motion. The neighbors of the

candidate motion are determined to compute the Euclidean distances between the set of training motions and a candidate motion. Then, K nearest neighbors are selected by ordering the distances in ascending order and selecting the first K motions. If K = 1, also known as 1-nearest neighbor, the predicted class of the candidate motion class is equal to the class of the training motion with the least Euclidean distance. If K $>$ 1, the predicted class of the candidate motion is equal to the most frequent class from among the classes of the K nearest neighbors of the candidate motion. Wu et al. [152] designed a recognition algorithm that uses the relative position of joints, as well as the angular velocity, and angular acceleration over joints over time as features. These features were combined with a KNN classifier, modified so that the voting system is dependent on the frames of the motion. Evaluation of the algorithm on the MSR-Action3D dataset and the UTD-MHAD dataset showed an accuracy of 92.34% and 90.47%, respectively.

Decision trees consist of a collection of nodes and edges connected to a tree structure constructed using the training motions [78]. Each node in the tree, also known as a split node, represents a test on one of the features, the edges represent the outcome of the test, and the leaf node represents the class labels. Each of the split nodes of a decision tree is regarded as a weak classifier since it returns an outcome based on the feature being used in the test. Hence, decision trees are referred to as an ensemble of weak classifiers [78]. The predicted class of a candidate motion is equal to the class in the leaf node that emerges from the traversal of the tree from its root node. An ensemble of decision trees in which a random subset of features is selected at each split node is called a "Random Forest" [78]. Garcia-Hernando and Kim [40] designed a recognition algorithm that relies on the joint position as features and a transition forest, which is an ensemble of decision trees that can discriminate the transitions between pairs of independent frames, as the classifier. Evaluation of the algorithm on the MSRC12 action dataset and the MSR-Action3D dataset showed an accuracy of 94.22% and 94.57%, respectively.

Hidden markov models (HMMs) are extensions of Markov chains and consist of several hidden states and state transitional probabilities showing the probability of moving between states [78]. The training motions belonging to a class label are used to build a hidden Markov chain for

that class label. The predicted class label of a candidate motion corresponds to the label of the Markov chain with the most likely state sequence. Xia et al. [153] designed a recognition algorithm that computes a histogram-based representation of joints tracked using the Kinect along with an HMM classifier. The authors evaluated the algorithm on the MSR-Action3D dataset and a self-collected 3D action dataset (not publicly available) comprising ten motions (e.g., walk and throw) performed by adults. On the MSR-Action3D dataset, the algorithm achieved a 98.61% accuracy when two-thirds of the motion data was used as training, while on the self-collected dataset, the algorithm achieved 90.21%.

Neural networks. Recently, researchers have started exploring algorithms that rely on deep learning to recognize human motions [153, 105]. Unlike traditional machine learning algorithms that require features that accurately reflect the discriminative information about the data [78], deep learning approaches do not need a careful selection of features for recognition [77]. These algorithms have been shown to achieve higher recognition accuracy compared to other recognition and are currently considered stateof-art algorithms for human motion recognition. Patsadu et al. [105] applied a Z-score normalization to joint positions and compared the recognition performance of a neural network classifier against an SVM and decision tree classifiers. The algorithms were evaluated on a dataset collected by the authors comprising three actions (sit down, lie down, and stand) performed by six adult subjects. Results showed that the neural network classifier achieved the highest accuracy (100%), followed by the SVM classifier (99.75%) and the decision tree classifier (93.91%). However, deep learning algorithms usually require a large amount of training data [77] and are more computationally expensive than template-based and traditional machine learning approaches, such as SVMs and KNNs.

Although all of these algorithms have been shown to achieve good recognition accuracy ($>$80%), they have only been evaluated on adults' motions. Children differ from adults in how they perform motions, as supported by our past work [64], so recognition systems trained on adults' motions will likely perform poorly on children's motions. As researchers and practitioners increasingly design motion-based applications for children that include motion recognition

algorithms, tailoring motion recognition systems to children's motion qualities to enable accurate recognition becomes important. However, the motion qualities that quantify the differences between how children and adults perform motions are not known, which makes tailoring motion recognition systems to children's qualities difficult. This dissertation work will focus on identifying the motion qualities that distinguish child motion from adults' motion to propose a set of guidelines for tailoring motion recognition systems to children's motion qualities.

2.4 Tailoring Stroke Gesture Recognition Systems to Children's Stroke Gestures

Prior work in stroke gesture recognition has found that stroke gesture recognizers have lower recognition accuracies for children's gestures compared to adults' gestures [6, 149]. Anthony et al. compared the recognition accuracy of stroke gestures produced by children ages 7 to 16 and adults using the $N-protractor stroke gesture recognizer and found that children's gestures had lower recognition rates (81%) compared to adults' gestures (90%) [6]. Similarly, Woodward et al. [149] found the recognition accuracy of gestures produced by children ages 5 to 10 (84%) was lower than those produced by adults (94%). Because of these findings, stroke gesture researchers have suggested the need to tailor stroke gesture recognizers to children's stroke gestures. Shaw et al. [121] investigated the differences between children's and adults' stroke gestures using the features defined by Anthony et al. [7] and Vatavu et al. [135]. Anthony et al. [7] identified 10 geometric features (e.g., path length) and two kinematic features (e.g., production time) that characterizes how both children and adults produce stroke gestures. Vatavu et al. [135] identified 12 relative accuracy features that characterize deviations from a reference gesture defined as a "gesture task axis" (e.g., shape deviation). Findings from their research showed differences between children and adults in several features. For example, the authors found that children exhibit longer path lengths and higher shape errors compared to adults. Furthermore, the authors found that children are more inconsistent in how they produce gestures as characterized by higher variations in some of the features, and this inconsistency causes higher recognition errors in children's gestures compared to adults' gestures.

Like stroke gestures, our past work showed that when two motion recognition systems that have been shown to achieve good recognition accuracy on adults motions (>80%) were tested on a dataset of children's and adults' motions, the systems performed worse for children's motions compared to adults' motions (DTW recognizer: adult = 76.3% [SD = 16.7%] child = 47.0% [SD = 19.9%], SVM recognizer: adult = 84.5% [SD = 16.6%] child = 60.9% [SD = 12.1%]), more details will be presented in Chapter 4. Hence, there is a need to tailor whole-body motion recognizers to children's motions. This work aims to address this issue by identifying the features that characterize the differences in how children and adults perform motions.

CHAPTER 3
SCOPE

In this chapter, we present the scope within which this dissertation work is situated by providing details about the children's agegroup we will focus on, the type of motion sensor used to track the motions, and the dataset we will use to analyze motion data.

3.1 Population

The ability to make movement relies on motor development, which is the study of the progression in a person's ability to perform motions [106]. Motor development is usually age-related [23] and although well-developed in adults, children are still in the process of developing their motor skills. Our work will focus on children ages 5 to 10 years old. The years 5 to 10 mark an important period in children's motor development [41]. At age 5, children are starting to make eager and confident movements [44] compared to children below age 5. Children at this age can also distinguish between the left and right sides of their body, also known as laterality [23]. By age 6, children become aware of the body, i.e., they understand what the body can do and can identify body parts. At age 7, children become aware of spatial and temporal orientations [41]. At age 8, children can understand body references [23], such as "swipe right." At age 8 to 9, children are more adept at perceiving motions, such that they can now track objects moving in an arc ([94] as cited in [23], p. 36) and at age 9, children are able to more accurately intercept moving objects [23]. Furthermore, children at this age are more skillful when performing motions [41]. At age 10, children have acquired expertise to identify body parts in the left and right side of the body and understand the relationship between body parts [54].

Through a review of child development, motor development, and motor learning literature, Cleland-Donnelly et al. [23] further asserted that the developmental characteristics of children in primary grades (kindergarten to grades 2 (K-2) and age range 5 to 7 years in the U.S.) differ from children in intermediate grades (grades 3 to 5 (3-5) and age range 8 to 10 years in the U.S.). For example, children in primary grades can perform fundamental movements, such as jumping, while children in intermediate grades can perform more complex movements, such as gymnastics. Therefore, children in the 5 to 10 age range are at varying levels of motor development and these variations could result in inconsistencies in how children perform motions. By focusing on this

age range, we can investigate how children at varying levels of development perform motions and how age impacts how children perform motions.

3.2 Motion Sensor

In this work, we use the Microsoft Kinect (Microsoft, n.d.), which is a motion sensor that has been commonly used to track how users move body parts during the performance of a motion for interaction purposes[22, 34]. The Microsoft Kinect is a low-cost tracking device that consists of an RGB camera, a depth sensor, and a microphone array (Figure 3-1A). The Kinect tracks the movement of a user skeleton containing 20 joints of the human body (Figure 3-1B) in 3-dimensional (3D) space: x for horizontal, y for vertical, and z for depth using a right-handed coordinate system. In this coordinate system, the positive direction of the x-axis is to the left of the Kinect, the positive direction of the y-axis is vertically upward, and the positive direction of the z-axis is away from the sensor (Figure 3-1C). The user skeleton is tracked at 30 frames per second, thus resulting in 3D skeletal frames that capture information about the posture of the body during the motion. The posture of the body is usually defined as the 3D position of each joint in the user skeleton expressed in meters.

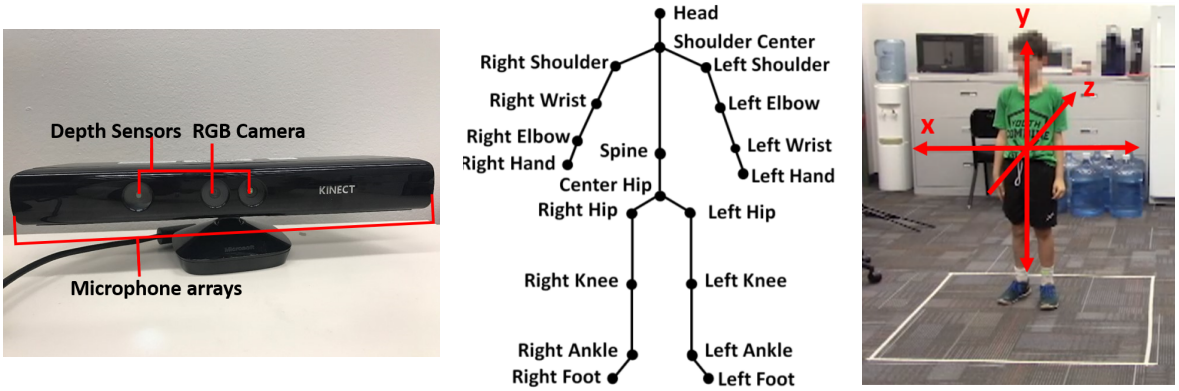


Figure 3-1. The Kinect Sensor and Skeleton. A) Kinect sensor (Photo courtesy of author). B) Kinect skeleton showing the 20 joints tracked. The skeleton also depicts a posture or pose in a motion. C) Coordinate frames centered at the Kinect (Photo courtesy of author. Source: Eakta Jain, Lisa Anthony, Aishat Aloba, Amanda Castonguay, Isabella Cuba, Alex Shaw,and Julia Woodward. 2016. Is the Motion of a Child Perceivably Different from the Motion of an Adult? ACM Transactions on Applied Perception 13, 4 (jul 2016), 1–17.).

3.3 Kinder-Gator Dataset

Publicly available human motion datasets usually include adults' motions (i.e., the motions are performed by adults; see table 2-1 in the previous chapter). When datasets have focused on children's motions, these motions have primarily focused on adults performing "childlike" motions (i.e., adults imitating how a child will perform the motion) [138]. This focus on adults is likely because recording children's movement using mocap devices (i.e., motion capturing) can be challenging as children sometimes struggle to follow instruction and have shorter attention spans [108]. Furthermore, there is additional cost and effort associated with recruiting children for research studies [31]. To the best of our knowledge, prior to our work, there was only one publicly available dataset that includes the motions of children and adults. Guerra-Filho and Biswas [46] created the Human Motion Database (HMD), which includes motions from 50 child and adult participants performing 70 motions. In this dataset, children performed the motion as demonstrated by the researchers (i.e., scripted motion) to maintain the consistency of the performance across the range of motions performed. However, to enable recognition of motions as users would naturally perform them in everyday contexts (e.g., during interaction with whole-body applications), it is important that our work focuses on "natural motions" to capture unique behaviors exhibited by children, as opposed to scripted motions, which may not capture these unique behaviors since children are mimicking an adult's performance of the motion. Because of the lack of publicly available datasets of children's natural motions, in our early work I helped collect the Kinder-Gator dataset [3] [1], a dataset of children's (ages 5 to 10) and adults' natural motions tracked using the Microsoft Kinect v1.

3.3.1 Dataset Collection

We collected a total of 58 motions in our dataset. A critical aspect for the creation of the dataset was to ensure it encompassed a diverse range of human motions, so motions in the dataset

[1] I acted as the first experimenter during the collection of half of the adult participants' motions during which I was responsible for informing the participant about the next motion to be performed and recording the motion as the motion is performed. I also led a team of students to organize the dataset for publication and led the effort for writing the paper, which was later accepted as a short paper to the European Graphics Conference (Aloba et al., Eurographics '18)

were chosen by reviewing prior studies involving whole-body motions [57, 63, 100]. Based on the review, we selected motions that were used in prior work, that both children and adults would be familiar with, and that we hypothesized would show differences between children and adults. We also included a set of simple "warm up" motions like waving to get participants familiar with the idea of performing motions and the study procedure. The motions collected from this review were classified into four categories:

- **Warm-up motions**. We selected nine motions that are easy to perform and are used in day-to-day activities (e.g., "Wave your hand").
- **Exercise motions**. We selected fourteen motions that induce exertion when performed and are commonly used in exercise and fitness activities (e.g., "Do five jumping jacks").
- **Mime motions**. We selected sixteen motions that involve the conceptualization of imaginary objects (e.g., "Climb an imaginary ladder").
- **Communication motions**. We selected nineteen motions are used to convey information (e.g., "Motion someone to stop").

Table 3-1 shows the full set of motions in the dataset.

3.3.2 Study Setup

Motions in the Kinder-Gator dataset were collected using Kinect v1.0 hardware and its accompanying Kinect for Windows SDK v1.8 software. Two researchers were responsible for prompting for the next motion to be performed and controlling the Kinect software. In each study session, a participant stood within an area denoted by a square (47 x 47 inches), forward-facing the Kinect and movements were only allowed within that specific area. The denotation of a specific area was to ensure that people did not move outside the tracking range of the Kinect. Participants performed all motions from a standing position. Before the start of each motion, participants stood with their arms outstretched in the form of a T-pose and then counted down from 3 to 1 while lowering their arms to their sides. The T-pose was required to get an accurate demarcation of the intended natural start of a motion.

The duration of each motion was dependent on the motion being performed. For "Wave your hand", "Walk in place", "Walk in place as fast as you can", "Run in place", "Run in place as

Table 3-1. A list of the 58 motions in the Kinder-Gator dataset

Warm-up	Exercise	Mime	Communication
Raise your hand	Put your hands on your hip and lean to the side	Push an imaginary button in front of you	Point at the camera
Raise your other hand	Put your hands on your hips and lean to the other side	Swipe across an imaginary screen in front of you	Motion someone to stop
Wave your hand	Put your hands on your hips and twist back and forth	Swipe across an imaginary screen in front of you with your other hand	Motion someone to come here
Wave your other hand	Touch your toes	Fly like a bird	Draw a [circle, square, triangle] in the air
Bow	Do a forward lunge	Fly like an airplane	Draw the letter [A, C, K, M, X] in the air
Raise your arm to one side	Lift your leg to one side	Swim	Make the letter [Y, M, C, A, K, P, T, X] with your body
Raise your other arm to the other side	Lift your other leg to the other side	Kick a ball	–
Bend your knee	Walk in place	Kick a ball as hard as you can	
Bend your other knee	Walk in place as fast as you can	Kick a ball with the other leg	–
–	Run in place	Kick a ball as hard as you can with that leg	–
–	Run in place as fast as you can	Throw a ball	–
–	Jump	Throw a ball as far as you can	–
–	Jump as high as you can	Throw a ball with your other arm	–
–	Do five jumping jacks	Throw a ball as far as you can with that arm	–
–	–	Punch	–
–	–	Climb an imaginary ladder	–

Table 3-2. Motions performed by participants in the Kinder-Gator dataset for which the motion was demonstrated to the participant by a researcher during the collection of the dataset

ID	Gender	Age	Hand	Grade Level
		Children		
337	F	5	B	Pre-K
595	M	5	B	Pre-K
106	M	6	R	K
290	M	6	R	K
342	F	6	R	K
474	F	6	R	1
169	M	8	R	2
103	F	8	R	3
723	M	8	R	3
644	F	9	R	4
		Adults		
565	F	19	R	High school
577	F	19	R	Some college
604	F	20	R	Some college
976	M	20	R	Some college
734	M	22	R	Undergrad
876	F	23	R	Undergrad
921	M	26	L	Undergrad
888	F	25	R	Grad
970	M	28	R	Grad
934	M	32	R	Grad

fast as you can", "Fly like a bird", "Swim", "Climb an imaginary ladder", and "Do five jumping jacks", the duration was typically about five cycles (10 steps or repetitions). For motions involving making poses with the body, the duration was 3 seconds, since the experiment staff required participants to hold the pose for 3 seconds. For all other motions, the duration of the motion varied depending on the participant. Participants always returned their hands down to their sides to demarcate the end of the motion. In each session, a participant performed one example of all 58 motions. In order to ensure that participants were performing the motions as naturally as possible, participants were allowed to perform the motion free form. That is, we did not require that the motions be performed in any predefined manner. When a participant did not understand how a motion was to be performed from verbal explanations alone, one of the researchers showed an example. (This occurred for four different children on two to six actions,

and two different adults on one to two actions: details in Table B-1 of the Appendix). To ensure the motions performed were natural, participants performed the motion only after the researcher had stopped demonstrating an example, so as to reduce the likelihood of imitation.

3.3.3 Participants

We recruited ten children and ten adults via flyers, emails, and advertisements on social platforms. Recruitment and study protocol procedures were approved by our Institutional Review Board. Child participant ages ranged from 5 to 9 (mean = 6.70, SD = 1.42). Five children were female. Two children were ambidextrous and none were left-handed. We focus on ages 5 to 9, since children in this age group are still growing in terms of their motor development, as detailed in the previous section. The adult participant ages ranged from 19 to 32 (mean = 23.40, SD = 4.33). Five adults were female and only one adult was left-handed. All participants were familiar with motion interaction systems such as the Microsoft Kinect. Participants each received a $10 gift card to a local grocery store as compensation. The full demographic information of participants in the dataset is shown in Table 3-2.

3.3.4 Data Collection

A total of 19 RGB videos and 1159 motion trials (58 motions x 20 participants) are included in the dataset; RGB video for all actions for one adult (ID: 934) and skeleton data for jump high for one adult (ID: 565) is missing due to a software error. The total time it took to perform the motions ranged from 247s to 363s (M = 301s, SD = 37.2) for children and from 302s to 424s (M = 344s, SD = 35.3) for adults. Each motion in the dataset is organized so that the first column has the timestamps. The timestamps are recorded in milliseconds and the difference between the last row and the second row of the first column gives the duration of the motion in milliseconds; the first row is the header. Subsequent columns have the x, y, and z positions of each joint as recorded by the Kinect. The timestamp is recorded in milliseconds and the joint positions are recorded in meters. The last two columns have the ID of the participant and the motion label; the values are always the same in each row for one motion (Figure 3-2). Then, each motion's data is saved as a .csv file.

Timestamp	HipCenterX	HipCenterY	HipCenterZ	...	FootRightX	FootRightY	FootRightZ	User-ID	Action
48546673.98	-0.0419986	-0.1512957	3.057717	...	0.0165785	-0.879243	2.962777	103	Raise-your-hand
48546708.98	-0.04312875	-0.151742	3.058163	...	0.0192515	-0.878694	2.960702	103	Raise-your-hand
48546740.99	-0.04372106	-0.1516228	3.058321	...	0.0170242	-0.878821	2.960905	103	Raise-your-hand
48546772.99	-0.0443055	-0.1512319	3.059047	...	0.0170806	-0.87882	2.960871	103	Raise-your-hand
48546808.99	-0.04423596	-0.1512616	3.059192	...	0.0176137	-0.878316	2.961083	103	Raise-your-hand
48546840.99	-0.04403488	-0.1516092	3.059681	...	0.0167917	-0.879489	2.961125	103	Raise-your-hand
48546873.99	-0.04366333	-0.1522055	3.060779	...	0.0174926	-0.877533	2.963976	103	Raise-your-hand
48546909	-0.04448649	-0.151911	3.061489	...	0.0174193	-0.877472	2.964743	103	Raise-your-hand
48546941	-0.04139955	-0.1517779	3.062791	...	5.94E-05	-0.910351	2.976981	103	Raise-your-hand
48546973	-0.04145111	-0.1520992	3.062808	...	0.0207837	-0.876891	2.962383	103	Raise-your-hand
48547009	-0.04190007	-0.1516161	3.062786	...	0.0183119	-0.881215	2.963866	103	Raise-your-hand
48547041	-0.04195235	-0.1515646	3.062788	...	0.0167279	-0.878802	2.964861	103	Raise-your-hand

Figure 3-2. Data format of motions in the Kinder-Gator dataset. The first column is the timestamp, subsequent columns are the x,y, and z positions of joints tracked by the Kinect sensor, and the last two columns are the ID of the participant and the motion being performed respectively. Each row is a pose/posture of the motion being performed.

3.4 Summary

To understand how children perform motions, in this dissertation, we will use the Kinder-Gator dataset, which focuses on children ages 5 to 10 performing motions tracked using the Kinect sensor. Specifically, we will show through several studies analyzing the motions in this dataset that there is a need to tailor whole-body motion recognition systems to children's motion qualities.

CHAPTER 4
ESTABLISHING A NEED TO UNDERSTAND CHILDREN'S MOTIONS

To motivate this work, we previously mentioned that children's motions are likely to differ from adults' motions due to differences in their body proportions and stages of neuromuscular development [59] and that because of this difference, whole-body recognition systems trained on adults' motions will likely perform poorly on children's motions. In this chapter, we present two studies to corroborate these ideas. Through these studies, we establish a need to understand the differences between how children and adults in order to tailor whole-body recognition systems to children's natural motion qualities.

4.1 Investigating People's Perception of Children's and Adults' Motions

To investigate if there are indeed differences between child and adult motion, in my first year, my co-authors and I conducted a perception study [1] to investigate whether naïve viewers can perceive the difference between child and adult motion when the motion is abstracted from all appearance cues (e.g., face, height, and build) [64].

4.1.1 Stimuli Preparation

We selected 6 dynamic motions from 8 participants (4 adults and 4 children) in the Kinder-Gator dataset [3]; at the time of this study, the Kinder-Gator dataset included motions from all child participants but only 4 adult participants, so we selected only 4 child participants to balance the data. The motions selected include "Do five jumping jacks", "Fly like a bird", "Jump as high as you can", "Run in place as fast as you can", "Walk in place", and "Wave your hand". Figure 4-1 illustrates the motion data, which has been pelvis-aligned to account for differences in joint angles and velocities of children and adults. For each of the motion, we created a point-light display (PLD) representation. A PLD refers to points of lights representing each joint in the human body 4-2. Since the Kinect tracks data at 30 fps, we later rendered the PLD representation as a video played back at 30 fps.

[1] I was responsible for abstracting the motion from all appearance cues (e.g., creating the point light display videos) and analyzing the data collected to extract the results. For the paper, I designed the box plots and scatter plots and wrote a section detailing how the motions were abstracted. This paper was later published in the Transactions of Applied Perception journal (Jain et al., TAP 2016)

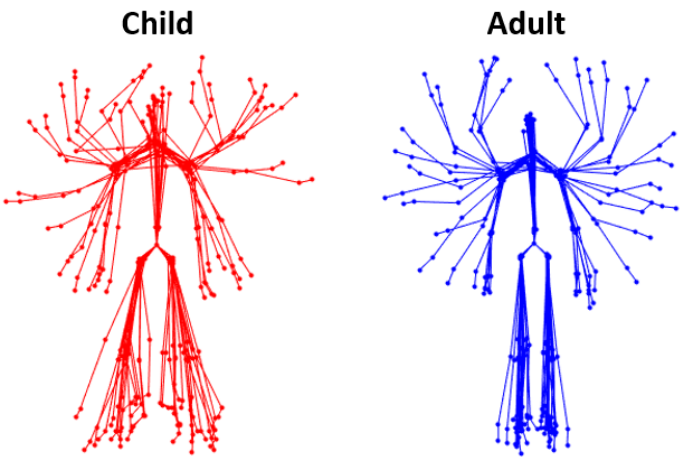


Figure 4-1. Motion data showing every tenth frame for jumping jacks for an adult motion. ©Jain et al. [64].

To account for natural size differences between children and adults, which can cue participants about the agegroup of the participant performing the motion, we scaled all stimuli videos to the same physical height. To achieve this scaling, we computed the mean value over all frames of the twenty joints tracked by the Kinect:

$$\mu_y = \frac{\sum_{i=1}^{N} \sum_{j=1}^{20} y_i^j}{20 * N} \tag{4-1}$$

Then, we subtracted the joint positions along the y dimension from its mean to center the positions relative to the mean. To scale the skeleton to canonical height within the video, we zoomed the render camera towards or away from the skeleton. Afterward, we added padding above and below to account for any extra space needed when the motion requires a larger amount of space to perform (e.g., "Jump as high as you can").

4.1.2 Results

To identify whether naive viewers are able to perceive the difference between children's and adults' motions, we created a survey wherein each question included a PLD video with an accompanying two-alternative forced choice question asking whether the motion belongs to a child or an adult and a free-form text answer about the motion that was performed. We recruited 34 participants but only 24 participants completed the survey.

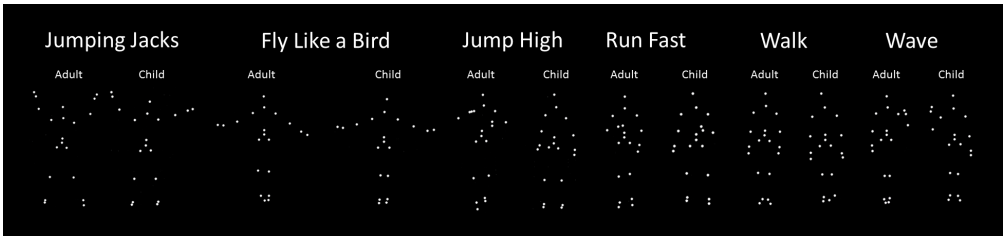


Figure 4-2. Point light display representations of motions considered in the perception study. Each joint is represented as a point of light (white dot).

For each participant, we computed the percentage of correct responses among all answers provided for the videos in the survey. We found that the percentage of correct responses for each participant in the study was above chance level (i.e., >50%; Figure 4-3A). A one-tailed t-test on percentage of correct response for both the child videos (mean = 62.1%, SD = 13.1%) and adult videos (mean = 70.0%, SD = 13.7%) videos evaluated independently showed significance above chance levels (child: $t = 4.23$, $df = 20$, $p < 0.05$, adult: $t = 6.72$, $df = 20$, $p < 0.05$; Figure 4-3B). We also found that that for all motions considered, the percentage of correct responses was significantly above chance (Figure 4-4).

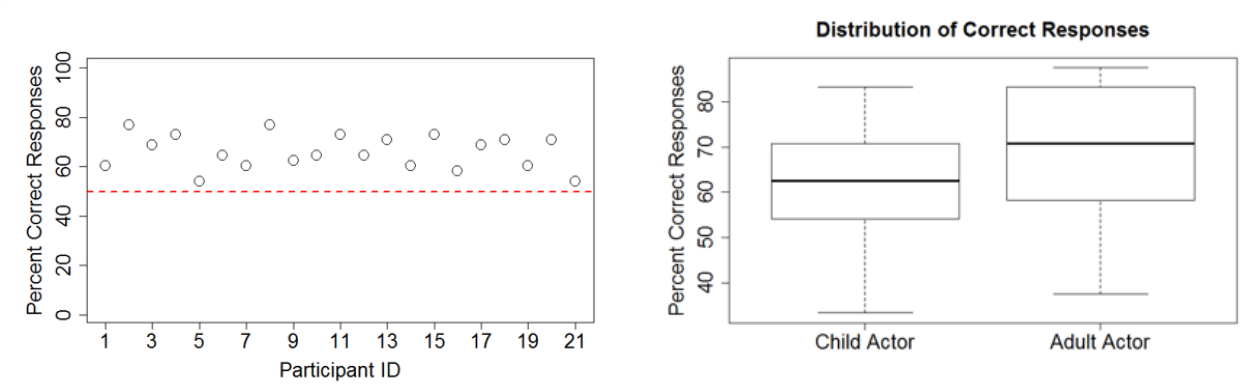


Figure 4-3. Results from analyzing the survey responses. A) Percentage correct response for all participants who responded to the survey. B) Boxplots showing median and interquartile ranges of the percentage correct response grouped by age group. ©Jain et al. [64].

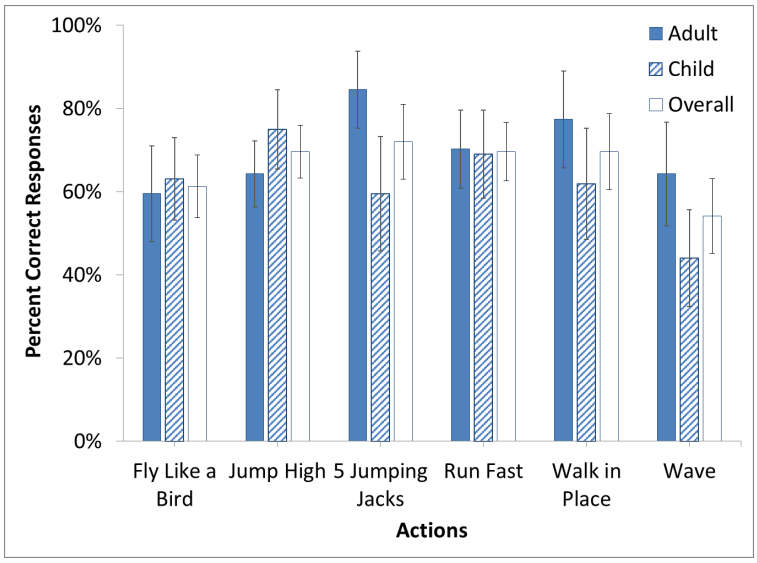


Figure 4-4. Percentage correct response by child and adult videos and by motions. Error bars are 95% confidence intervals. ©Jain et al. [64].

4.1.3 Discussion

From our results, we saw that naïve viewers can perceive the difference between children's and adults' motions at levels significantly above chance and about 70% for dynamic motions, such as walking and running. These findings establish that children move differently from adults. Therefore, whole-body recognition systems trained on adults' motions will likely perform poorly on children's motions.

4.2 Evaluating the Performance of Skeleton-Based Motion Recognition Systems

To investigate the performance of whole-body recognition systems trained on adults' motions on children's motions, we selected three motion recognition algorithms that have achieved good recognition accuracy (>80%) on adults' motions [19, 22, 5]. These algorithms include:

1. A template-based adaptation of stroke gesture recognizers to multiple articulation paths
2. A template-based recognizer using Dynamic Time Warping (DTW)
3. A machine-learning recognizer using Support Vector Machines (SVMs)

For each motion recognition algorithm, we compared its recognition performance on children's and adults' motions by training and testing the recognizer on children's motions, and in a separate experiment, training and testing the recognizer on adults' motions. We used a subset of motions from the Kinder-Gator dataset [3] for the comparison. The motions were selected to be distinct to prevent conflicts, which can negatively affect the performance of the recognizer on both children's and adults' motion. Prior to using these motions to test the algorithms, we manually excluded the T-pose from each of the motion (we mentioned in the previous chapter that during the collection of the dataset, each participant was required to perform a T-Pose prior to performing the motion to get an accurate demarcation of the start of the motion). In this section, we detail the motion selection process and the recognition process.

4.2.1 Motion Selection

Motion recognition algorithms are usually trained on motion datasets. Motions in these datasets are usually selected to be distinct from each other since conflicts among motions can negatively impact recognition. However, as shown in Table 3-2 in the previous chapter, motions in the Kinder-Gator dataset are not distinct. We chose a representative (distinct) set of gestures from the dataset by removing gestures that are currently out of scope of our work and gestures with obvious conflicts in the dataset. The gestures we considered out of scope of our work are 3D stroke gestures, in which the emphasis is on the shape or symbol being drawn (e.g., "Draw the letter A in the air") and periodic gestures: gestures in which the same set of poses occur multiple times (e.g., "Run in place"). We excluded periodic gestures because instances of the same gesture type that have different numbers of repetitions might not be matched properly to each other. After exclusion, we grouped gestures that are similar in terms of how they are performed. First, we grouped gestures that have mirrors, since the gestures being performed are the same, just with the opposite limb (e.g., "Raise your other hand" is a mirror of "Raise your hand"). We also grouped gestures that are the same motion differing only in strength (e.g., "Throw a ball" vs. "Throw a ball as far as you can"; "Kick a ball" vs. "Kick a ball as hard as you can") as these motions will be articulated the same way. Lastly, we grouped the gestures ("Point at the camera", "Motion

someone to stop", and "Push an imaginary button in front of you"). These gestures are difficult to distinguish using the Kinect v1 alone, because their differences are based on the position of the finger, which this sensor cannot track. We selected one gesture from each such group for the representative set. Our final representative set comprises 14 gestures (Table 4-1). The rest of this dissertation work will focus on the motions in this representative set.

4.2.2 Adaptation of Template Gesture Recognizers to Multiple Articulation Paths

Prior work in stroke gesture research has designed 2D stroke gesture recognition systems for classification of 2D stroke gestures (e.g., letters and symbols) [147, 69]. For example, Wobbrock et al. [147] designed the $1 recognizer, a template-based recognizer for identifying unistroke gestures. Anthony et al. [8] and Vatavu et al. [134] designed $N and $P respectively as extensions to $1 for recognition of multistroke gestures. Li [76] also designed protractor and Anthony et al. [9] designed $N-protractor as closed-form solutions to $1 and $N respectively. These recognizers have been shown to accurately recognize 2D stroke gestures, $1, $N, and $P achieve x,y, and z accuracies when tested on 2D stroke gesture datasets respectively. Both 2D stroke gestures and motions are similar in that they both involve paths moving in space over time: the former involves the path of the finger on the touchscreen surface while for the latter, each body part moved by a user creates a path in 3D space over the duration of the motion. Therefore, 2D stroke gesture recognition systems can be adapted for recognition of 3D gestures. Kratz and Rohs [69] designed the $3 gesture recognizer, a template-based recognizer that extends $1 to enable recognition of 3D gestures. The researchers also designed Protractor3D, a closed-form solution that extends

Table 4-1. Our distinct set of 14 gestures from the Kinder-Gator dataset

Gestures	
Touch your toes (TYT)	Do a forward lunge (DFL)
Point at the camera (PAC)	Lift your leg to one side (LYL)
Raise your hand (RYH)	Jump (J)
Raise your arm to one side (RAS)	Kick a ball as hard as you can (KBH)
Bend your knee (BYK)	Throw a ball as far as you can (TBF)
Put your hands on your hip and lean to one side (PHL)	Swipe across an imaginary screen in front of you (SIF)
Punch (P)	Bow (B)

protractor to 3D. However, these 3D gesture recognizers are limited to single articulation paths (i.e., one 3D point in each frame over time). However, motions tracked by motion sensors usually have multiple articulation paths (i.e., more than one 3D point in each frame over time). For example, motions tracked by the Kinect v1, such as those in the Kinder-Gator dataset [3] have an articulation path for each of the 20 joints tracked by the Kinect. Hence, these algorithms will need to be adapted to multiple articulation paths to enable accurate recognition of motions.

To compare the recognition performance of these algorithms for children's and adults' motions, we developed a 3D template-based gesture recognition system that adapts the $3 recognizer [69] to multiple articulation paths by modifying the algorithm and using components from protractor3D [70] and $P [135]. Because stroke gesture recognition algorithms are trained and tested on adults' gestures and prior work has shown that these algorithms perform poorly for children's gestures compared to adults' gestures [6], we posit that our adaptation will also reflect these findings. That is, our algorithm will perform poorer for children's motions compared to adults' motions. In this section, we discuss the steps for adapting the algorithms to multiple articulation paths and our recognition experiment using the representative set of motions from the Kinder-Gator dataset [3].

4.2.2.1 Normalization

Given a set of M points that define an articulation path of a gesture, first, we smooth the articulation path (i.e., the path traveled by a joint during movement) of each of the joints using an exponential moving average filter with alpha = 0.1. Prior work [124] has found that this filter can remove noise from joint data without introducing any smoothing artifacts to the data. Figure 4-5 shows the articulation path of the right elbow joint before and after smoothing the path. Then, we use the $3 resampling step [69], which selects points so that the articulation path is defined by N equidistant points. We resampled the articulation path of a joint to 32 points, which has been shown to be adequate for recognition of stroke gestures [147]. After resampling, $3 rotates the articulation path so that its indicative angle (i.e., the angle between the first point and the centroid point of the articulation path [147, 69]) is zero. After rotation, $3 scales the articulation path

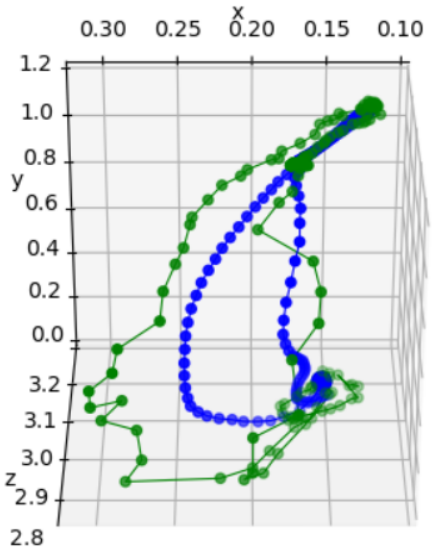


Figure 4-5. Articulation paths when raising the right elbow before (green) and after (blue) smoothing using the exponential moving average filter.

non-uniformly to a reference cube of size 100^3 dimension. However, non-uniform scales are not effective when the range of points is close to zero [147]. For example, in Figure 4-6, the range of points of the green articulation path along the x-axis is close to zero. Hence, we used uniform scaling as proposed in the $P recognizer [135], which applies the same scaling factor to all dimensions. The uniform scaling factor is equal to the maximum range of points from among the ranges of points from each dimension for that gesture instance. After scaling, $3 translates the articulation path so that its centroid is at the origin. The above processes ensure that similar gesture instances that differ only by speed, rotation, size, and position respectively can be matched to each other. Lastly, $3 finds the optimal alignment between two articulation paths, which is the alignment that gives the minimum average Euclidean distance [147, 69]. We use the closed form solution in Protractor3D [70] to find the optimal alignment. We rotate with respect to the base orientation of the motion using the approach from the original Protractor gesture recognizer [76], so that dissimilar gesture instances that differ only in their orientation can still be distinguished. For example, the gestures “Raise your arm to one side” and “Raise your hand”, differ only in terms of their orientation (horizontal 90^o and vertical 90^o). The normalization step improves recognition accuracy in the face of minor gesture articulation variations by users. Figure

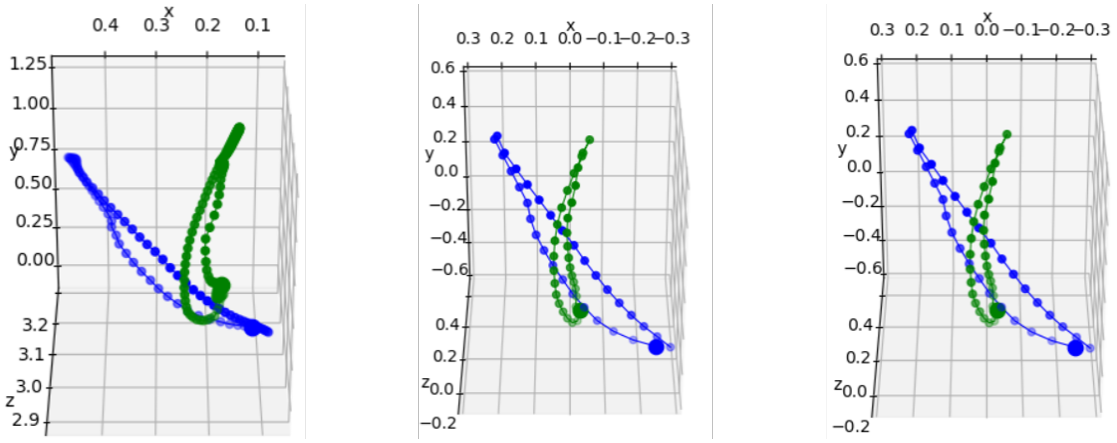


Figure 4-6. Motion paths for smoothed gestures “Raise your arm to one side” (blue) and “Raise your hand” (green) for the same joint before and after normalization. A) Motion paths for the gestures “Raise your arm to one side” (blue) and “Raise your hand” (green) for the same joint. B) Same paths prior to optimal alignment. C) Same paths after optimal alignment, where global rotational orientation is maintained.

4-6 shows example motion paths for one joint in two motion instances after applying all the normalization steps.

4.2.2.2 Recognition

After normalization, $3 consecutively matches the points of the articulation path of the test gesture to be recognized to each gesture in the training set. The gesture in the training set whose articulation path has the lowest Euclidean distance to that of the test gesture is selected. To extend this approach to multiple articulation paths, for each test gesture C and a gesture t in the training set T, we normalize each joint j^i in C and t. Thus, C becomes $C^{'}$ and t becomes $t^{'}$. Then, we compute the optimal alignment between joint $j^i_{C^{'}}$ in $C^{'}$ and the corresponding joint $j^i_{t^{'}}$ in $t^{'}$ and compute the average Euclidean distances e_1 and e_2 after rotating $j^i_{C^{'}}$ in $C^{'}$ to match $j^i_{t^{'}}$ in $t^{'}(r_c)$ and $j^i_{t^{'}}$ in $t^{'}$ to match $j^i_{C^{'}}$ in $C^{'}(r_t)$ respectively:

$$e_1 = \frac{\sum_{k=1}^{N} \sqrt{\sum_{q\in\{x,y,z\}} (r_c(k)_q - t^{'}(k)_q)^2}}{N} \tag{4-2}$$

$$e_2 = \frac{\sum_{k=1}^{N} \sqrt{\sum_{q\in\{x,y,z\}} (C^{'}(k)_q - r_t(k)_q)^2}}{N} \tag{4-3}$$

$$d_i = min(e_1, e_2, e_3, e_4) \tag{4-4}$$

where N is the number of points being considered. To account for directional differences resulting from making the same motions with different limbs (e.g., raising the left hand vs. raising the right hand), we rotated j^i in $C^{'}$ 180^o ($flip\_j^i$) by negating the position along the x-dimension. We computed the distances e_3 and e_4 after rotating $flip\_j^i$ to j^i in $t^{'}$ ($r\_flip\_j^i$) and rotating j^i in $t^{'}$ to $flip\_j^i$ (r_t_flip), respectively. Then, we compute d_i as the minimum of e_1, e_2, e_3, and e_4. The minimum average Euclidean distance d defines how close the gesture path between C and t is, and is calculated as the sum of d_i over all joints p in C, where p is the number of joints tracked by the motion sensor (i.e., p=20 for the Kinect). The gesture from T with the lowest Euclidean distance to C is the recognition result, that is, t for which $d = \sum_{i=1}^{p} d_i$ is minimum.

4.2.2.3 Testing

We tested the recognizer on children's and adults' motions in two separate experiments. First, we trained the recognizer on children's motions and tested on children's motions. Then, in a separate experiment, we trained the recognizer on adults' motions and tested on adults' motions. For testing, we use a leave-one-out cross validation (LOOCV) method [148]. In LOOCV, gestures from one participant is used for testing while all other participants' gestures are used for training. The training/testing process is repeated until the recognizer has been tested on gestures from all participants. Because the Kinder-Gator dataset [3] includes 10 children and 10 adults, for each of the children's and adults' recognition experiments, we select motions from 9 participants for training and the remaining participant's motions is left out for testing. We repeat training/testing 10 times so each participant's motions are used in one trial for testing. The accuracy of each trial is the number of correctly classified gestures from among the gestures selected for testing, and the overall accuracy is the sum of the accuracies across all trials divided by the number of trials (i.e., 10 trials).

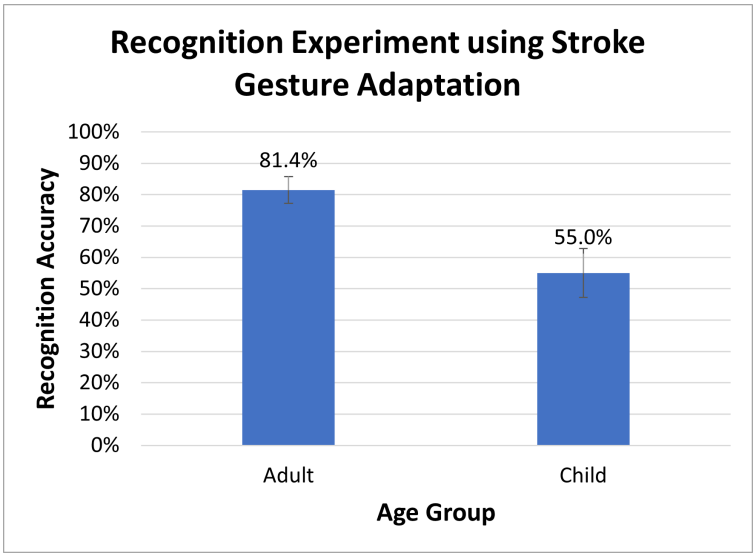


Figure 4-7. Recognition performance of our adaptation of template-based gesture recognizers to multiple articulation paths on children's and adults' motions in the Kinder-Gator dataset. Error bars indicate 95% confidence interval.

4.2.2.4 Results

When evaluated on the 14 gestures from the representative set of children's and adults' motions in the Kinder-Gator dataset, our gesture recognizer achieved a recognition accuracy of 81.4% (SD = 6.9%) on adults' motions and recognition accuracy of 55.0% (SD = 12.6%) on children's motions (Figure 4-7).

4.2.3 Template-Based Recognition Experiment Using DTW

For the next recognition experiment, we use Dynamic Time Warping (DTW). As a reminder, DTW measures the similarity between time sequence of motion measurements or features extracted from the motion measurements (e.g., angle vectors of each body part in a skeleton) [19].To prepare the motions for recognition using DTW, first, we smooth the articulation path using the same exponential moving average filter we used in the previous section and then pre-process articulation paths of the motions using the $3 normalization steps described in the previous section. This preprocessing step is important to ensure that that differences in terms of speed, rotation, size, and position for instances of the same gesture type can be removed, which will improve recognition accuracy in the face of minor gesture articulation variations by

users. However, because DTW already performs a time warping to align motions during the matching step [19], the algorithm is not affected by differences in the speed at which the motion was performed. Hence, we excluded the resampling step in $1 [147], which normalizes motions in terms of speed. We applied the following normalization steps:

1. Rotate the articulation path so that its indicative angle (i.e., the angle between the first point and the centroid point of the articulation path) is zero.

2. Scale the articulation path uniformly to a reference cube.

3. Translate the articulation path, so its centroid is at the origin

Like the previous experiment, we use the same testing approach (i.e., we trained and tested the recognizer on children's motions and in a seperate experiment, trained and tested the recognizer on adults' motions using a leave-one-out cross validation (LOOCV) method [148]). When evaluated on the 14 gestures from the representative set of children's and adults' motions in the Kinder-Gator dataset, the Dynamic Time Warping (DTW) recognizer achieved a recognition accuracy of 73.6% (SD = 14.3%) on adults' motions and recognition accuracy of 60.7% (SD = 15.2%) on children's motions (Figure 4-8).

4.2.4 Machine Learning Recognition Experiment Using SVM

For this recognition experiment, we used an SVM recognizer designed by Cippitelli et al. [22]. The researchers showed that their recognizer achieved high recognition accuracies on multiple publicly available datasets of adults' motions. For example, the recognizer achieved accuracies up to 86.1% on the Florence3D dataset and up to 95.1% on the UTKinect dataset (see Table 2-1 in Chapter 2 for more details about these datasets). To prepare the motions for SVM classification, given one instance of a motion, we used the following steps defined by the authors [22] to extract features that capture distinct information about motion types:

Posture feature extraction. For each skeleton frame of the motion, a feature vector representing each posture is created. To create the feature vector, each joint is first translated with respect to the hip joint so that the centroid of its path is at the origin. The goal of this translation is

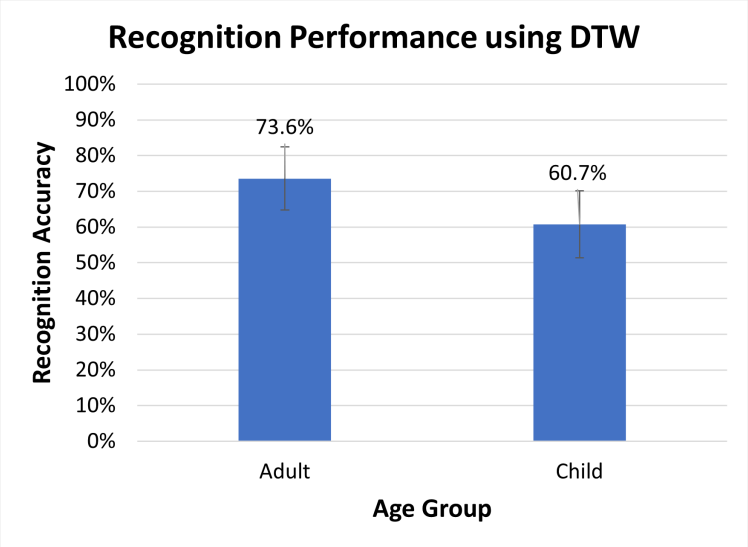


Figure 4-8. Recognition performance of a DTW recognition algorithm on children's and adults' motions in the Kinder-Gator dataset. Error bars indicate 95% confidence interval.

to account for differences in the position where the motion is performed. Then, the joint is normalized using the Euclidean distance between the shoulder center and hip joints to make the resulting feature invariant to the height and build of the person.

Postures selection. Once the posture feature vectors have been created for all frames, similar postures are grouped using a K-means clustering algorithm [53], where K for this algorithm can be as low as 3 and as high as the minimum number of frames among the motions being considered. K-means will partition the feature vectors such that each vector within a cluster is closer to the mean for that cluster, also known as the cluster center, compared to the mean of any other cluster. The K-means algorithm can be summarized as follows:

1. Select K vectors from the feature vectors randomly and assign them as cluster centers for each of the K clusters.
2. For each cluster, compute the Euclidean distance between each feature vector and its cluster center.
3. Assign each feature vector to the cluster with the minimum Euclidean distance.
4. Calculate the cluster center of each of the K cluster as the average of all the vectors in that cluster.
5. Repeat the second, third, and fourth steps until there are no changes in the cluster in which the vectors are assigned.

Because the vectors to be clustered are time series consisting of 20 3D joints tracked using the Kinect sensor, each cluster center is a vector consisting of 60 values. The cluster centers [C_1, C_2, ..., C_K] are selected as the features, where the number of cluster centers is dependent on the size of K.

Activity features computation: The K-means algorithm assigns a label to each posture that corresponds to the cluster the posture was grouped into. The labels are arranged in the order that the postures are recorded in the motion data, and this arrangement is used to concatenate the cluster centers. For example, if the motion being considered has five postures [P_1, P_2, P_3, P_4, P_5] with corresponding cluster labels [2,3,2,1,2] based on the results from the clustering algorithm, the cluster centers are concatenated in the order [C_2, C_3, C_1]. The concatenated cluster centers are used to represent the motion.

As mentioned earlier, the goal of these steps (posture feature extraction, postures selection, and activity features computation) is to extract features (i.e., the concatenated cluster centers) that capture discriminative information about the motion being performed. We applied these steps on each of the motions in the representative set of motions we selected from the Kinder-Gator dataset [3] to extract features. Cippitelli et al. [22] noted that their algorithm performs better with a high number of clusters, so we chose K = the minimum number of frames from among all 14 motions being considered and across all participants in the age group whose motions are being recognized (i.e., child or adult). For adults' motions, K = 46 while for children's motions, K = 47. Therefore, adults' motions are represented by a feature vector of 2760 (i.e., K*60) values and children's motions are represented as a feature vector of 2820 values.

4.2.4.1 Testing

After representing all the motions using feature vectors representing the concatenated cluster centers, Cippitelli et al. [22] classified the motions using a Support Vector Machine (SVM). As a reminder, an SVM classifier uses an optimal hyperplane to separate the training motion in higher dimensions to create a decision boundary for classification [77]. Using the same

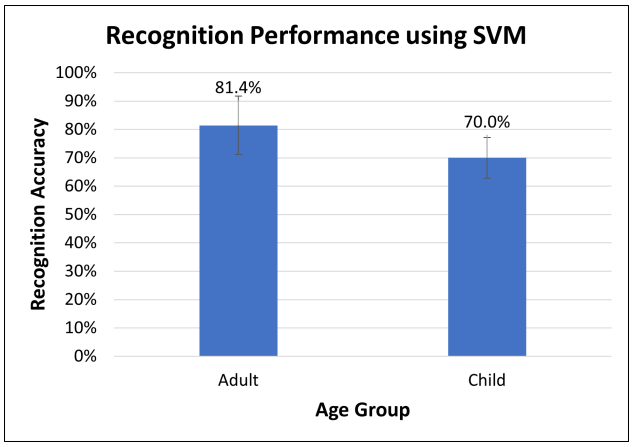


Figure 4-9. Recognition performance of Cippitelli's recognition algorithm on children and adults' motions in the Kinder-Gator dataset. Error bars indicate 95% confidence interval.

approach as the previous two recognition algorithms, we tested the performance of the algorithm [22] by training and testing on children's motions and in a separate experiment, training and testing on adults' motions using the leave-one-out cross validation (LOOCV) method [148].

4.2.4.2 Results

For adults' motions in the Kinder-Gator dataset [3], the algorithm achieved a recognition accuracy of 81.4% (SD = 16.6%) , which is similar to results found by the researchers on publicly available datasets for adults' motions. For children's motions, the algorithm achieved a recognition accuracy of 70.0% (11.6%) (Figure 4-9).

4.3 Summary

We evaluated three motion recognition algorithms that have been shown to achieve high accuracy on adults' motions. For each recognition algorithm, we compared its recognition performance on children's versus adults' motions by training and testing the recognizer on children's motions and in a separate experiment, we trained and tested the recognizer on adults' motions. Our findings showed that both the template-based and machine-learning recognizers we tested achieved lower recognition accuracies on children's motions compared to adults' motions (Our template-based Adaptation: adult = 81.4% [SD = 6.9%] child = 55.0%[SD = 12.6%], DTW recognizer: adult = 73.6% [SD = 14.3%] child = 60.7% [15.2%], SVM recognizer: adult = 81.4% [16.6%] child = 70.0% [11.6%]). These findings provide evidence to show that motion

recognition algorithms that have been shown to achieve high recognition accuracy on adults' motions will perform poorly on children's motions. Taken together, the findings presented in this chapter establish that there are motion qualities present in children's motions that distinguish them from adults' motions. However, these motion qualities are not known, which makes tailoring motion recognizers to children's motions difficult. In the following chapters, we will focus on characterizing children's natural motion qualities by investigating the differences between children's and adults' motions. This investigation will inform an understanding of how children move, which we can then use to propose guidelines that future research can use to tailor motion recognition systems to children's motion qualities to enable accurate recognition.

CHAPTER 5
TOWARD UNDERSTANDING CHILD AND ADULT MOTION ARTICULATIONS

Our perception study discussed in the previous chapter showed that there are perceivable cues that differentiate child motion from adult motion. However, what exactly is being perceived has not been identified: what are the differences between child and adult motion? If we can identify these differences, then we can characterize children's natural motion qualities. To do this, we focus on understanding the nuances in how users articulate motions, which we propose can be achieved by a) identifying the joints that are critical to performing motions to investigate variations in how children and adults move their body parts when performing motions and b) identifying features that can quantify how the motion is produced to aid in the analysis of child and adult motion. In this chapter, we present two studies to help in characterizing the differences between child and adult motion. In the first study, we present a method that identifies the joints that are critical to performing motions. In the second study, we present a preliminary study that quantifies differences between children's and adults' motion.

5.1 FilterJoint: Toward an Understanding of Whole-Body Motion Articulation

To understand the nuances in how children and adults articulate whole-body motions, we designed a method that facilitates an investigation of the variations in how users move body parts as they perform whole-body motions [1]. Our method, which we call filterJoint, selects the key joints that are actively moving during the performance of the motion. Whole-body motions tracked using motions sensors, such as the Kinect [3] are characterized by multiple articulation paths, each defined as the 3D positions of joints over the duration of the motion. However, not all joints tracked by the motion sensor are necessary to perform a motion. Tracking errors of the motion sensor [19] and unintentional movements from the user could make joints that are not supposed to move appear to move. For example, to raise one's hands, only the joints in the upper limbs (e.g., hand, elbow) should move, while any movement in the joints of the lower limbs (e.g., foot) is likely due either to tracking noise of the sensor or to unintentional movements from the user (see Figure 5-1). Including these articulation paths during the analysis of whole-body motion

[1] (Aloba et al., ICMI 2020). I was responsible for designing the filterJoint method, and analyzing the results using a template-based stroke gesture adaptation. I was also the first author on the publication, responsible for writing our research findings.

articulation could result in incorrect inferences about how users perform whole-body motions as well as introduce noise in the recognition process, thus affecting recognition accuracy. The paths along which these joints move in space over time can then be analyzed to make inferences about how children and adults articulate whole-body motions.

5.1.1 The FilterJoint Method

To identify the joints that are necessary to articulate a wholebody gesture, we designed a method that automatically identifies only the joint paths that are due to intentional movements from the user. In this method, which we call the filterJoint method, we attempt to select the actively moving joints for a particular whole-body motion gesture instance by computing the variations in joint movement and using a k-means algorithm to group the variations into two clusters. The joints in the cluster with the higher mean (i.e., higher range of motion) are selected as the actively moving joints for that motion. Given a motion instance m for which each joint is defined by N $3D$ points:

$$m = \begin{bmatrix} j_1^1 & j_1^2 & \cdots & j_1^K \\ j_2^1 & j_2^2 & \cdots & j_2^K \\ \vdots & \vdots & \vdots & \vdots \\ j_N^1 & j_N^2 & \cdots & j_N^K \end{bmatrix}$$

where j_i^K is a 3D point showing the position of j^K at time instance i, N is the number of points in the motion path of j^K , and K is the number of joints tracked by the sensor. We apply the following steps to extract the moving joints for each motion instance:

First, we smooth the articulation path of each of the joints in m using an exponential moving average filter with alpha = 0.1. Prior work [20] has found that this filter can remove noise from joint data without introducing any smoothing artifacts to the data. The aim of the smoothing process is to reduce articulation path noise due to tracking issues of the motion sensor.

We compute variations for each joint movement j^K in m by computing the standard deviation (SD):

$$SD = \sqrt{\frac{\sum_{i=1}^{N}(j_x^k - j_{ix}^{\bar{k}})^2 + (j_y^k - j_{iy}^{\bar{k}})^2 + (j_z^k - j_{iz}^{\bar{k}})^2}{N}}$$

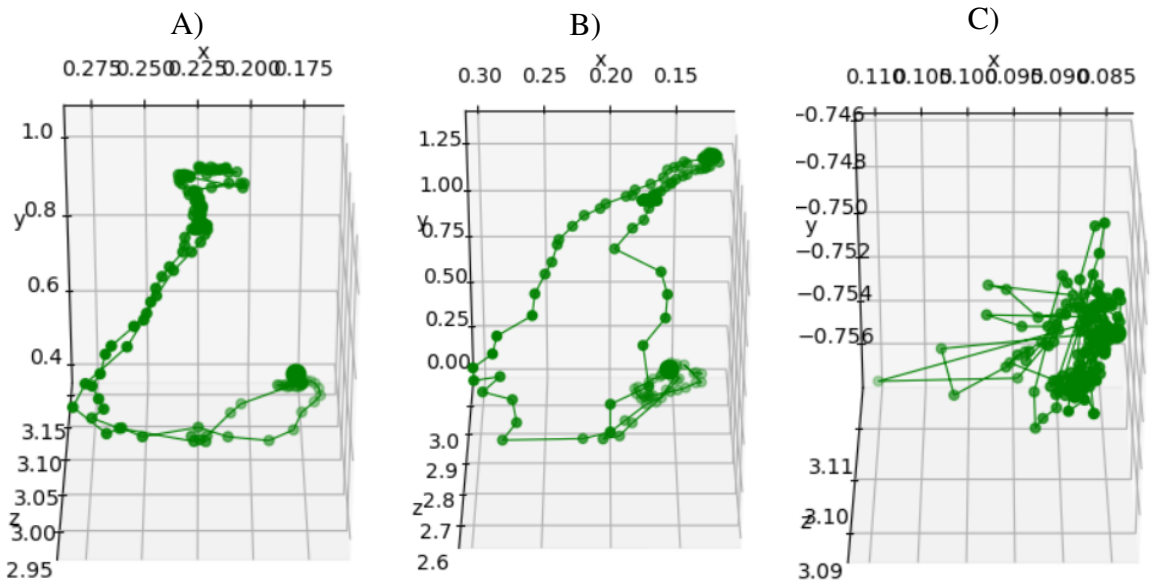


Figure 5-1. Motion paths when raising the right hand. A) Joints of the right hand and B) right elbow that move. C) Joint of the foot that does not move.

where $SD(m) = [SD(j^1), \ldots, SD(j^K)]$. In an ideal scenario, we expect that joints that do not move should have standard deviations close to zero. However, due to tracking noise exhibited by the motion sensor and unconscious joint movements as users articulate motions, what happens is that joints that should not move actually do move.

Thus, we attempt to classify intentionally moving joints using k-means [8], a clustering algorithm that partitions a set of values into k disjoint clusters, such that each value within a cluster is closer to the mean for that cluster compared to the mean of any other cluster. We posit that $SD(m)$ can be partitioned into two clusters: joints that move intentionally and joints that should not move. We set the cluster size to two because we expect that joints that users are intentionally moving should have much higher variations compared to joints that move due to noise, such that these variations can be grouped into two separate clusters. Hence, we used k-means to partition $SD(m)$ with the number of clusters set to two ($k = 2$). The k-means algorithm will define two clusters c_1 and c_2, each having a mean x_1 and x_2, respectively, and guarantees that one of the means is greater than the other. We set the initial means of the clusters as $max(SD(m))$ and $min(SD(m))$ and k-means updates the means iteratively until all the values have been

allocated to a cluster. The cluster with the higher mean (e.g., higher range of motion), c_1, we select as containing all the intentionally moving joints.

For whole-body gestures that require users to move all of their body parts, for example, when making a "Jump", we expect that most of the joint variations will be fairly close to each other. However, with our 2-cluster k-means approach, these joint variations will be forced into two clusters and the cluster c_1 with the higher mean will be selected even though we would not necessarily agree that the joints in c1 move more than the joints in c_2. To account for such motions, we use a threshold, such that if the difference between the cluster means is below the threshold, then it means that the cluster with the lower mean (i.e., c_2) still includes joints that are actively moving. We compute the threshold as the average of the absolute difference between x_1 and x_2 for all motions being considered. If the absolute difference between x_1 and x_2 for motion m is less than the threshold, we repeat step 3 on c_2 (i.e., the cluster with the lower mean) to further partition the joints. Joints that were initially clustered into c_2 clearly do not move with as much variation in position (i.e., as actively) as the joints previously clustered into c_1. On the other hand, splitting c_2 again should be able to separate joints that have even less variation (i.e., due to tracking noise from the motion sensor and unintentional movements from the user) from more active joints. We add the resulting joints in the new cluster with the higher mean to the previously selected set of actively moving joints (c_1 from step 3).

The motion m is now defined by the joints selected by the filterJoint method. For example, for a "Raise your hand" motion (right hand), our method will select the right hand, right wrist, and right elbow from the full set of 20 joints (Table 5-1).

Table 5-1. Example motion types in the Kinder-Gator dataset with joints selected by our filterJoint method.

Gestures	Joints Selected
Raise your hand	Right hand, Right wrist, Right elbow
Bend your knee	Right knee, Right ankle, Right foot
Point at the camera	Right hand, Right wrist, Right elbow

5.1.2 Evaluating the FilterJoint Method

To evaluate that the filterJoint method does indeed select the key actively moving joints, we adapt template-based gesture recognition algorithms. As mentioned in Chapter 2, these algorithms compare an articulation of a gesture to other articulations of gestures using point-by-point correspondence in order to select the articulation that most closely resembles the one being tested based on a distance metric. Since the goal of our work is to understand how users articulate whole-body motions, we use template-based approaches as opposed to state of the art machine learning approaches because the former can be used to make intuitive inferences about whole-body gesture articulation (e.g., whether users show variations in how they articulate instances of the same gesture type). In template-based gesture methods, changes in the articulation of the path will result in changes in the point-by-point correspondence that affect recognition accuracy in a predictable way. In contrast, state-of-the-art machine learning approaches, such as LSTMs and HMMs, use complex models and a "black-box" approach [8], which makes it less clear how changes in the articulation path affect recognition results.

5.1.2.1 Evaluation

We compared the recognition accuracy when using our filterJoint method compared to a baseline method (i.e., involving all joints tracked by the motion sensor). In template-based algorithms, recognition accuracy is an indicator of the similarity between articulation of instances of the same gesture type. That is, the higher the recognition accuracy, the higher the consistency between users' articulations of instances of whole-body gestures of the same type. Joint articulation paths resulting from tracking noise of the sensor and unintentional movements from the user will result in variations in how users articulate similar gestures, which will introduce noise during the recognition process. This noise will detract from a recognizer being able to accurately distinguish between motion types, so exclusion of such joints should improve recognition accuracy. Since our filterJoint method attempts to filter out such joint articulation paths, a higher recognition over the baseline method will indicate that our method was successful in removing noisy or unimportant joints. For our evaluation, we use adults' whole-body gestures

from the publicly-available Kinder-Gator dataset [3]. We focus on only adults for this evaluation because stroke gesture research has shown that adults tend to be more consistent than children in how they articulate gestures. Hence, template-based stroke gesture recognizers perform better on adults' gestures compared to children's gestures [2]. We used our representative set of 14 gestures for the evaluation.

For the evaluation, we used our adaptation of 3D template-based gesture recognizers to multiple articulation paths described in Chapter 4. As a reminder, we adapted the $3 gesture recognizer [69] to multiple articulation paths by modifying the normalization steps using components from protractor3D [70] and $P [135] as well as updating the recognition step to account for multiple articulation paths. To apply this algorithm for our evaluation, depending on the method being used (filterJoint vs. baseline method), the recognition step will only use joints selected by the filterJoint method and all joints tracked by the motion sensor respectively. Because the joints selected by the filterJoint method can vary in number depending on how a user performs the motion, in the recognition step, the filterJoint method is first applied to the candidate motion to select the actively moving joints and then the articulation paths of these joints are compared to the corresponding articulation paths of each of the template motion in the template set. For testing, we use a leave-one-out cross validation (LOOCV) method [148].

5.1.2.2 Results

Our filterJoint method achieved a recognition accuracy of 90.7% [SD = 6.8%] while the baseline method achieved a recognition accuracy of 81.4% [SD = 6.9%] (Figure 5-2). A pairwise t-test showed that the filterJoint method was significantly more accurate than the baseline ($t(9) = 6.09$, $p < 0.01$). Therefore, we can conclude that the filterJoint method successfully filters out noisy joints that are not important to articulation.

5.1.3 Discussion

We have presented the filterJoint method, which filters out noisy or unimportant joint motion paths in whole-body gesture articulations, because including such joints during the analysis of whole-body gesture articulation could result in incorrect inferences about how users

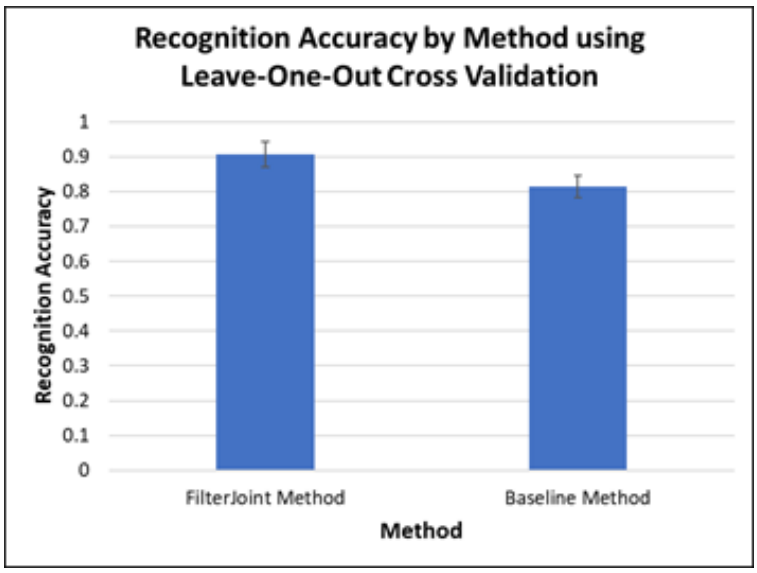


Figure 5-2. Performance of the gesture recognition algorithm. Error bars indicate 95% confidence interval.

make motions. The articulation paths of joints selected by our method (i.e., key joints that are necessary to articulating the whole-body gestures), can be analyzed to investigate nuances in how users articulate whole-body gestures. We present two case studies using the Kinder-Gator dataset [3] that showcase the kinds of new insights about how users articulate whole-body gestures that our filterJoint method enables, and we highlight implications from these insights for selecting whole-body gesture sets. While the first case study focuses on showing the efficacy of our method in showing how users perform motions, our second case study focuses on shows how our method can be used to differentiate child and adult motion. While we do not present general findings about whole-body gesture articulation across all possible gesture types, we demonstrate the applicability of our filterJoint method as a tool that designers can use to understand how users articulate different gesture types, and help them select better gesture sets.

5.1.3.1 Case study: identifying overlaps in motions

Although we removed motions with obvious conflicts when selecting our original motion set, our recognition results still showed recognition errors. We analyzed the confusion matrix, which is a tool for exploring data in recognition algorithms, to get a better understanding of the

confusions between motions (Figure 5-3). The specific confusions the recognizer makes among motions may reveal insights about whole-body gesture articulation patterns.

An analysis of the confusion matrix from our LOOCV recognition experiment showed that there are whole-body gestures that are confused for each other even though they do not share any similarities that we would expect in terms of their poses. Since we have identified the actively moving joints using our filterJoint method, we can examine the articulation paths of these joints to understand how participants within the dataset articulate whole-body gestures. Our adaptation allows us to explicitly determine how the articulation of whole-body gestures contributes to confusions between otherwise dissimilar motions. For example, the worst confusions occurred between gestures "Throw a ball as far as you can (tbf)" (7) and "Punch (p)" (6) (Figure 5-3, row TBF, column P and row P, column TBF). We believe that these confusions are likely because both gestures involve the act of swinging the arm(s) forward, so there is a high chance that users will articulate these gestures in a similar fashion. Supporting our expectations, we found that for 6 out of 10 participants, the actively moving joints selected by our filterJoint method for the p gesture overlapped with those selected for the tbf gesture for at least one other participant. This finding suggests an overlap between participants' articulation of the p and tbf gestures. As another example, the gesture "Kick a ball as hard as you can (kbh)" (8) was confused as "Do a forward lunge (dfl)" (1) (Figure 5-3, row KBH, column DFL) and "Throw a ball as far as you can" (1) (Figure 5-3, row KBH, column TBF). From our understanding of how the gestures kbh and dfl are likely to be articulated, we expect that kbh should only involve movement of joints in the lower limb (e.g., knee and foot) while the dfl motion should involve movement of both upper and lower limbs. Contrary to our expectation, based on the joints selected by the filterJoint method, we found that 8 out of 10 participants actively moved their upper limbs during the articulation of the kbh motion. Although this behavior is not expected, prior work in biomechanics has shown that upper limb movements can help to maintain a balance when only one foot is on the ground [67, 145], which occurs when articulating the kbh motion.

	Bend your knee (BYK)	Bow (B)	Do a for-ward lunge (DFL)	Jump (J)	Kick a ball as hard as you can (KBH)	Lift your leg to one side (LYL)	Point at the cam-era (PAC)	Punch (P)	Put your hands on your hips and lean to the side (PHL)	Raise your arm to one side (RAS)	Raise your hand (RYH)	Swipe across an imag-i-nary screen in front of you (SIF)	Throw a ball as far as you can (TBF)	Touch your toes (TYT)
BYK	9	0	1	0	0	0	0	0	0	0	0	0	0	0
B	0	9	0	0	0	0	0	0	0	0	0	0	0	1
DFL	0	0	10	0	0	0	0	0	0	0	0	0	0	0
J	0	0	0	10	0	0	0	0	0	0	0	0	0	0
KBH	0	0	1	0	8	0	0	0	0	0	0	0	1	0
LYL	0	0	0	0	0	10	0	0	0	0	0	0	0	0
PAC	0	0	0	0	0	0	10	0	0	0	0	0	0	0
P	0	0	1	0	1	0	0	6	0	0	0	0	2	0
PHL	0	0	0	0	0	1	0	0	9	0	0	0	0	0
RAS	0	0	0	0	0	0	0	0	0	10	0	0	0	0
RYH	0	0	0	0	0	0	0	0	0	0	10	0	0	0
SIF	0	0	0	1	0	0	0	0	0	0	0	9	0	0
TBF	0	0	0	0	0	0	0	3	0	0	0	0	7	0
TYT	0	0	0	0	0	0	0	0	0	0	0	0	0	10

Figure 5-3. Confusion matrix for the recognition results with 9 training templates. Rows represent the frequency of times the motion was categorized as the column. Correct recognitions are along the diagonal

The gesture “Put your hands on your hips and lean to the side (phl)” (9) was also confused with one other gesture: “Lift your leg to one side (lyl)” (1) even though the phl motion does not in and of itself share a similarity with lyl (Figure 5-3, row PHL, column LYL). From our understanding, we expect that the phl gesture should not actively involve lower limb movements (e.g., foot). Our analysis of the actively moving joints corroborates our expectations; however, we found that the participant whose phl gesture was misclassified did actively move their foot during articulation. This foot movement could have resulted in the participant’s articulation of the phl motion being confused for another participant’s articulation of the lyl motion because there would

have been an overlap between the joints selected by our filterJoint method. We also expected that the lyl motion should not involve joints in the upper limb (e.g., hand), but we found that four participants actively moved their hand or shoulder during the articulation of this gesture. Prior work has noted that the farther a user leans to one side, the higher the chance that the user will lose balance due to a shift in their center of mass and gravity [103]. To compensate for the shift, people may raise the leg opposite of the way they are leaning [97]. Hence, the participant may have raised their leg or moved their upper limbs to maintain balance during the articulation of the phl and lyl gestures respectively, thus, affecting their motion articulations.

The above findings suggest that when motions require balance, users may intentionally move additional joints to maintain balance that we did not initially expect would be critical to the movement. It is a positive outcome that our filterJoint approach is robust enough to ensure that these joints are not filtered out. These findings also suggest that there are between-user inconsistencies in whole-body gesture articulation, so a designer who might have done what we did to select gestures would still see conflicts. Our approach can be used to detect these conflicts, which can make it easier for designers to select a better set of gestures. For example, based on our findings, we might want to exclude the gestures "Punch (p)" and "Kick a ball as hard as you can (kbh)" from our gesture set to avoid the conflicts arising from users' articulation of these gestures with other gestures like "Throw a ball as far as you can (tbf)" and "Do a forward lunge (dfl)", respectively. We recommend excluding p and kbh since these gestures had more variations in how users articulated them compared to the gestures they conflicted with (Figure 5-3).

We recommend that for applications that require a unique set of whole-body gestures (e.g., exergames), designers should consider applying our filterJoint method after selecting a distinct set of gestures to further exclude gestures that overlap due to nuances in users' whole-body gesture articulation.

5.1.3.2 Case study: understanding motion articulations

In this case study, we use the filterJoint method to understand children's and adults' motion articulations. The Kinder-Gator dataset [3] includes both children's and adults' motions, so we

also investigated how the articulation paths selected by our filterJoint method can show differences in children's and adults' whole-body gesture articulations. We investigated children's and adults' degree of agreement for each of the gesture types in our representative set (Figure 5-4). We defined the degree of agreement as equal to the total number of unique joint combinations selected within a gesture type. For example, for the gesture "Raise your arm to one side", adults in the Kinder-Gator dataset have only one unique combination ("hand right + wrist right" (10 adults)). Children have four unique combinations for the same motion ("hand right + wrist right" (2), "hand right + wrist right + elbow right" (6), "hand right + hand left + wrist right + wrist left + elbow right" (1), and "hand right + hand left + wrist right + wrist left + elbow right + shoulder right + foot left + knee right + head" (1)). We conducted a paired samples t-test to compare degree of agreement in children and adults across gesture types and found a significant difference ($t(13) = 3.38; p < 0.01$). Adults had a higher level of agreement ($mean = 5.43 \pm 2.71$) than children ($mean = 7.14 \pm 2.21$).

These findings suggest that children are less consistent than adults in how they articulate whole-body gestures, which has important implications for the selection of whole-body gesture sets. For example, gesture sets that are suitable for adults are not necessarily suitable for children. Using the Kinder-Gator dataset, we found that children had the highest degree of agreement for gestures that use only arm movements (e.g., "Raise your hand", 2 unique combinations and "Raise your arm to one side", (4)). On the other hand, children had low degree of agreement for gestures involving the whole-body (e.g., "Jump", 9 unique combinations) and gestures involving lower-limb movements (e.g., "Lift your leg to one side", 7 unique combinations) (Figure 5-4). This finding could be due to the increased coordination among many joints required to perform more complex movements. Since children are still developing their motor abilities [23, 54], they are less likely to have experience coordinating multiple joints to perform movements compared to adults. In addition, the higher degree of agreement in arm motions, which only involves movement of the upper limbs, can be attributed to balance and postural stability. The lower-limb motions in the Kinder-Gator dataset (e.g., "Lift your leg to one side") usually require that the user

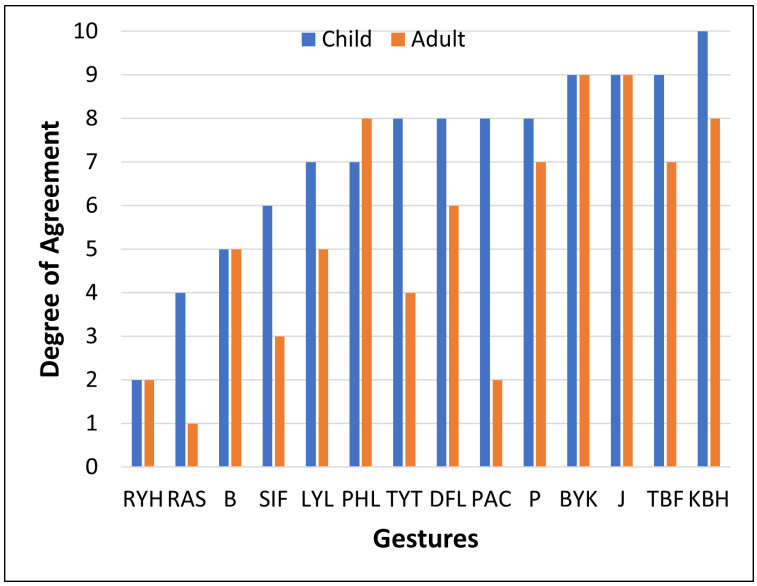


Figure 5-4. Degree of agreement of children and adults for gestures in our gesture set (lower = higher agreement).

maintains balance when performing the movement on one leg. Children are more likely than adults to move other joints to maintain balance since they are still developing their postural stability [55] (e.g., the arms play a functionally relevant role in balance among children [55]).

Hence, we recommend that, if possible, designers of whole-body gesture applications for children should prefer gestures that require only upper limb movements, especially arm movements, since children will articulate those motions more consistently, and thus recognition will be more accurate.

5.2 Quantifying Differences Between Child and Adult Motion Using Gait Features

To inform an understanding of children's natural motion qualities, for example, the reason behind the inconsistencies in how children perform motions discussed in the previous section, we focus on identifying features that can quantify the differences between child and adult motion[2]. We initially concentrated on walking and running motions since findings from the perception study (presented in the previous chapter) showed that naïve viewers could perceive the difference

[2]I was responsible for reviewing the literature to identify the gait features and analyzing the features on walking and running motions. I also led the effort on the paper, responsible for writing and presenting our research findings, which was later accepted as an invited paper in the International Conference on Human-Computer Interaction (Aloba et al., HCII 2019).

between child and adult motion for dynamic motions (e.g., walking and running) with about 70% accuracy [64]. To analyze these motions, we relied on the gait literature, which has identified features that characterize walking and running motions from a physiological perspective [128, 39, 60].

5.2.1 Gait Analysis

Gait is defined as one's manner or style of walking [36]. The analysis of gait is defined as the systematic study of human locomotion [39, 126], using the cycles and steps in the motion (Figure 5-1). A gait cycle (stride) is defined as the period between a foot contact on the ground to the next contact of the same foot on the ground again [39, 60]. A gait step is defined as the period between a foot contact on the ground to the next contact of the opposite foot, also known as half a gait cycle [60].

Wilheim and Eduard Weber [142] pioneered the study of spatial and temporal gait features by showing that human locomotion can be measured quantitatively. This finding led to the development of different quantitative methods for analyzing gait kinematics, of which the most commonly used is the placement of 3D markers along segments of the human body [13]. Since then, gait kinematics have been studied extensively. Researchers [25, 39, 60, 85, 112, 118, 123] have placed reflective markers on adults walking at different speeds, and have analyzed distance and time features of their gait such as stride length, step length, walk ratio, stride time, cadence, and speed. Gait analysis has also been used to identify individuals from their gait. Gianaria et al. [42] achieved 96% accuracy on classifying adults by gait by extracting gait features from Kinect data and feeding the features into a support vector machine (SVM). Prior work has also analyzed features from children's gait [10, 38, 35, 122]. Dusing and Thorpe [35] analyzed the cadence of children ages 1 to 10 walking at a self-selected pace, and found that cadence reduces as age increases. Barreira et al. [10] also studied the cadence of children walking freely in their environment. They found that children spent more time at lower cadences (0–79 steps/minute) compared to cadences signifying moderate or vigorous physical intensity (120 steps/min).

A limitation of the studies reviewed above is that they mainly focus on either children or adults. Some prior work has studied the comparison between child and adult motion, but they either focus on very young children [28] or older children [101] rather than a range of younger and older children. Davis [28] extracted and fed features collected from children's and adults' gait into a two-class linear perceptron to differentiate between the walking motion of young children (ages 3 to 5) and adults. He found that gait features can be used to differentiate between young children's and adults' walking patterns with about 93–95% accuracy. Oberg et al. [101] also compared gait features across ages from 10 to 79 years, and found that the speed of the gait and length of a step reduces with age. In this study, we extracted gait features such as cadence, step time, and step length from walking and running motions of children in the Kinder-Gator dataset [3] (i.e., ages 5 to 9) and adults to quantify the differences between children's and adults' walking and running motions.

5.2.2 Gait Features

We surveyed the literature on gait analysis [25, 28, 39, 60, 85, 91, 112, 118, 128], and identified ten features commonly used to characterize a person's gait. Gait analysis has been historically utilized to analyze walking or running motion that involves moving a distance away from the starting point. One feature that is commonly examined in gait analysis is the step length. The step length measures the distance between feet along the direction of motion, which for moving motions is parallel to the floor. However, the walking and running motions in the Kinder-Gator dataset [3] involve moving in place instead of moving away from a starting point. Therefore, the direction of motion is perpendicular to the floor instead of parallel to the floor, so we calculated the perpendicular distance (i.e., step height). This adaptation from step length to step height is valid because both measure the peak distance between feet.

Of the ten features we identified, we eliminated cycle length, which is the distance between successive placements of the same foot (measured as two step lengths). We eliminated this feature because, for in-place motions, the same foot returns to nearly the same location between steps. Measuring successive placements of the same foot using two step heights instead of two step

lengths would imply the participant is moving continuously upward, e.g., climbing up a ladder, rather than the in-place motions in our dataset [3]. We categorized the remaining nine gait features into spatial and temporal feature groups (Figure 5-5). Spatial features are distance-based; they include features which are dependent on the length (height) of a step in their computation. Temporal features are time-based. They include features which are dependent on time in their computation. We chose these nine features because prior research has shown that they are unique per person [25], that is, analogous to a fingerprint, and can be used as a biometric measure [14]. The nine features include:

Step width (m). This is a spatial feature. The step width is the maximum lateral distance between feet during a step [128]. It is measured as the horizontal distance between the position of one foot and the other foot during a step. This feature evaluates how wide or narrow the step taken is.

Step height (m). This is a spatial feature and is an adaptation of the step length. Step length is defined as the distance by which a foot moves in front of the opposite foot [128, 39, 60]. Since the walking and running motions in the Kinder-Gator dataset [3] involve moving in place rather than forward over a distance, we define the step height as the distance a foot travels above the other foot during a step. It is measured as how high above the ground vertically a foot is during the highest part of a step.

Relative step height. This is a spatial feature, and is defined as the length of a step in relation to the height of the person [28]. It measures the ratio between the step length (we use step height because they are walking in place) and height of the person. The relative step height is an important feature to consider as it normalizes the step height by the person's height, hence eliminating differences in step heights due to variations in height across people (e.g., children and adults).

Walk ratio (m/Steps/Minute). This is a spatial feature. It is an index used to characterize a person's walking pattern, and is measured as the ratio between the step length (we use step height) and the cadence (rate at which a person walks) [112, 118]. This feature is relevant to

dynamic motions, as previous research notes that it reflects participants' balance and coordination when performing a motion [8, 40].

Step time (s). This is a temporal feature which defines the time duration of a step [128]. It can further be defined as the time it takes a foot to complete one step. It is measured as the duration from when the foot leaves the ground to the time when the foot touches the ground again in completion of a step.

Cycle time (s). This is a temporal feature and is also known as stride time. The cycle time can be defined as the time it takes a foot to complete one cycle (two steps). It can be measured as the time between two consecutive steps of the same foot along the horizontal (we use vertical) trajectory [28].

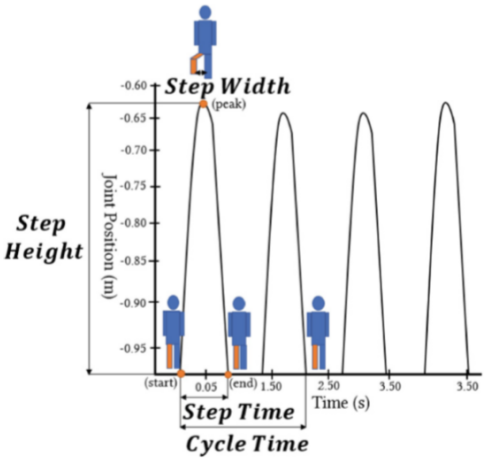
Cycle frequency (1/s). This is a temporal feature and is also known as stride frequency. It is defined as the number of cycles per unit time and can be computed as the inverse of the cycle time [27, 28, 85]. Prior research shows that participants' preferred cycle frequency optimizes energy cost [85].

Step Speed (m/s). This is a temporal feature, and is defined as the ratio between the step length (we use step height) and the step time [128]. It defines how fast a step is completed and can help in understanding the pace of a motion.

Cadence (steps/min). This is a temporal feature and is defined as the rate at which a person walks. It is measured as the number of steps taken per minute [25, 39, 85, 91], and reflects the level of energy being exerted.

5.2.3 Analysis of Motions

We used walking and running motions from the Kinder-Gator dataset [3], namely: "Walk in place" (walk), "Walk in place as fast as you can" (walk fast), "Run in place" (run), and "Run in place as fast as you can" (run fast) for the analysis. Recall that this dataset includes the motions of 10 adults and 10 children. To compute the gait features for each person-motion pair, we depend on knowledge of the stance phase (when the foot is on the ground [25]) and swing phase (when the foot is away from the ground [25]). Hence, we needed to identify the frames of each motion



$$\textbf{Relative Step Height} = \frac{\textit{Step Height}}{\textit{Participant Height}}$$

$$\textbf{Walk Ratio} = \frac{\textit{Step Height}}{\textit{Cadence}}$$

$$\textbf{Cycle Frequency} = \frac{1}{\textit{Cycle Time}}$$

$$\textbf{Step Speed} = \frac{\textit{Step Height}}{\textit{Step Time}}$$

$$\textbf{Cadence} = \frac{\textit{Number of Steps}}{\textit{Total Time Duration}}$$

Figure 5-5. Formulas for the nine gait features extracted from the data

that corresponded to the step boundaries. We manually extracted these frames from the point-light display videos using a video annotation toolkit called EASEL [139]. Two researchers annotated subsets of the point-light display videos of the Kinect data for all of the motions. To ensure balanced labeling, the videos were counterbalanced between each annotator by age group (child, adult) and motion. Also, for similar motions (run & run fast, walk & walk fast), the same annotator annotated the same participant for both motions. For each video to be annotated, we created three tracks in EASEL. Frames for the start (foot is on the ground), peak (foot is at its maximum position), and end (foot is returned to the ground) were recorded on the first, second, and third track respectively. The frames start-peak-end are the frames within a step. Previous research [39] has suggested that the analysis of gait can be done with either the foot, knee, hip, or pelvis joint, so we used the left foot joint from the Kinect skeleton tracking data in our analysis. Once all the frames had been annotated, we exported the annotation session, which creates an output CSV file with all the frames and the corresponding tracks recorded. We used this file for feature computation based on the start-peak-end frames.

We automated the feature computation process by extracting the corresponding foot positions and time stamps from the data for the frames we had manually extracted. The foot positions and time stamps were used to compute the gait features. For features involving

computations per step or cycle, we averaged the values over the total number of steps or cycles in that motion. Therefore, each participant has one data point per motion for each gait feature. A two-way repeated measures ANOVA was used to analyze the main effect of age group and motion type and the interaction effect between them. Whenever we found no interaction effect between age group and motion, we recomputed the two-way repeated measures ANOVA without the interaction effect in the model, and report that. For features where we found a significant effect of motion, we conducted a Tukey post-hoc test to identify motion pairs that are significantly different. We present results for all of the gait features we considered in our study. All means and standard deviations for features in the analysis can be found in Table 5-2, and they are expressed in units commonly used in the analysis of gait [11].

5.2.4 Results–Spatial Features

Distance-based features generally showed no significant effect of age group; hence, we conclude that they cannot be used to distinguish adult and child motion. However, these features show a significant effect of motion type, which serves to validate our approach of using features from the analysis of gait, despite the differences in motion structure (i.e., in-place motions versus moving along a distance).

Step width. Recall that the width of a step is computed as the horizontal distance between both feet during a step. A two-way repeated measures ANOVA on step width with a between-subjects factor of age group (child, adult), and a within-subjects factor of motion (walk, walk fast, run, run fast) found no significant effect of age group ($F_{1,18} = 12.15$, *n.s.*). The lateral placement of the feet for adults is roughly the same as that for children. This similarity may be because both adults and children have less control over the horizontal distance between their feet since in-place motions involve vertical movements. However, we found a significant main effect of motion ($F_{3,57} = 50.47$, $p < 0.0001$). Post-hoc tests only identified a difference between walk and run fast ($p < 0.001$). People have wider step widths when running fast than when walking (see Table 5-2), irrespective of the age group. Bauby & Kuo [11] asserted that wider steps are an

Table 5-2. Mean (SD) of Gait Features by Age Group and Gestures. (*) denotes significant effect at $p < 0.05$

	Age Group			Motions				
Gait Feature	Child	Adult	Sig.	Walk	Walk Fast	Run	Run Fast	Sig.
Step Width (m)	0.16 (0.04)	0.17 (0.06)		0.14 (0.05)	0.16 (0.05)	0.16 (0.05)	0.18 (0.04)	*
Step Height (m)	0.10 (0.06)	0.10 (0.10)		0.10 (0.09)	0.08 (0.08)	0.09 (0.08)	0.13 (0.09)	*
Relative Step Height	0.09 (0.05)	0.06 (0.06)		0.07 (0.07)	0.06 (0.06)	0.07 (0.05)	0.10 (0.06)	*
Walk Ratio (m/steps/min)	0.00063 (0.00078)	0.00070 (0.00061)		0.0010 (0.0010)	0.00051 (0.00057)	0.00048 (0.00044)	0.00061 (0.00051)	*
Step Time (s)	0.33 (0.11)	0.43 (0.15)	*	0.53 (0.14)	0.37 (0.12)	0.32 (0.08)	0.29 (0.05)	*
Cycle Time (s)	1.05 (0.47)	1.26 (0.49)	*	1.84 (0.30)	1.05 (0.34)	0.97 (0.25)	0.78 (0.14)	*
Cycle Frequency (1/s)	1.15 (0.44)	0.90 (0.32)	*	0.56 (0.09)	1.07 (0.37)	1.14 (0.35)	1.34 (0.25)	*
Step Speed (m/s)	0.32 (0.17)	0.25 (0.27)		0.17 (0.14)	0.24 (0.20)	0.28 (0.20)	0.44 (0.27)	*
Cadence (steps/min)	203 (79)	166 (61)	*	99 (16)	193 (66)	203 (56)	243 (48)	*

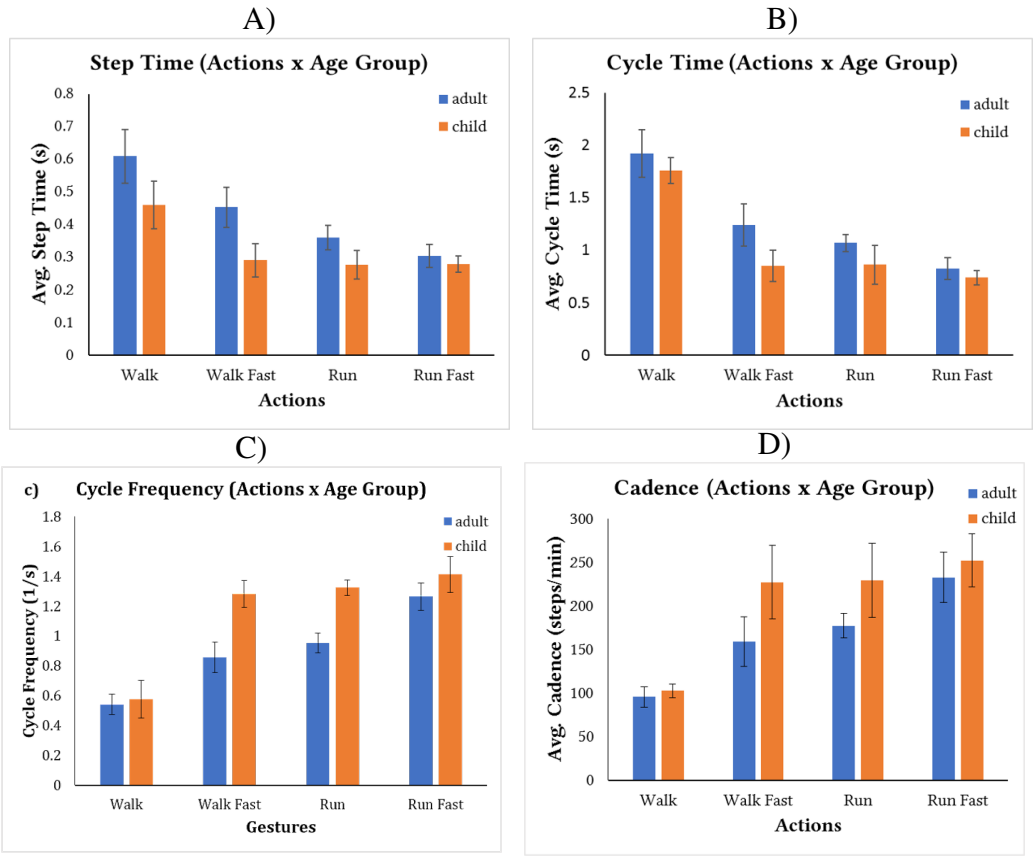


Figure 5-6. Effect of Motions and Age Group on selected Gait Features. Error bars indicate 95% confidence interval. A) Step Time. B) Cycle Time. C) Cycle Frequency. D) Cadence

advantage in stability, thus, participants widen their steps when running fast compared to walking to improve coordination. No difference in step width was found between other pairs of motions.

Step height. The height of a step is computed as the vertical distance between when the foot is on the ground, and when the foot is at its maximum position (peak). We focused on the left foot step heights because we annotated the left foot joints. We computed the ground for each step as the minimum between the (start) and (end) position. A two-way repeated measures ANOVA on step height with a between-subjects factor of age group (child, adult), and a within-subjects factor of motion (walk, walk fast, run, run fast) found no significant effect of age group ($F_{1,18} = 0.002$, *n.s.*). Our analysis found that the average step height for children and adults is the same (see Table 5-2). This finding is surprising given the typical difference in height between children and adults. The range helps to illuminate what is really happening. The range of step heights for

children was (min: 0.01m, max: 0.25m, med: 0.11m), and adults (min: 0.008m, max: 0.40m, med: 0.06m): thus, adults can raise their feet higher than children but children tend to raise them proportionally higher on average. We also found a significant effect of motion ($F_{3,57} = 5.03$, $p < 0.0001$). Post-hoc tests showed the following motion pairs differed: run/run fast ($p < 0.05$). Participants have higher step heights when running fast compared to the other motions (see Table 5-1): this confirms their exertion level was higher when running fast during our study.

Relative step height. The relative step height is computed as the ratio of the step height to the height of the person performing the motion. We estimated the height of a participant using the difference between the head and the foot along the vertical dimension (y axis). A two-way repeated measures ANOVA on relative step height with a between-subjects factor of age group (child, adult), and a within-subjects factor of motion (walk, walk fast, run, run fast) found no significant effect of age group ($F_{1,18} = 1.51$, *n.s.*). The large variance in the relative step height in children and adults may be the reason why we found no significant difference (Table 5-2). However, adults generally have a lower average relative step height compared to children. This finding is expected since their average step heights were the same, but adults are taller than children. We also found a significant effect of motion ($F_{3,57} = 5.03$, $p < 0.05$). Post-hoc tests showed that the following motion pairs differed: run/run fast ($p < 0.05$). Like step height, children and adults have a higher relative step height when running fast compared to just running (see Table 5-1).

Walk ratio. The walk ratio, a measure of balance and coordination, is computed as the ratio of the step height and the cadence (a temporal feature). A two-way repeated measures ANOVA on walk ratio with a between-subjects factor of age group (child, adult), and a within-subjects factor of motion (walk, walk fast, run, run fast) found no significant effect of age group ($F_{1,18} = 12.15$, *n.s.*). The average walk ratio for children and adults can be found in Table 5-2. The similarity in walk ratio between children and adults may be because the walking pattern is influenced by how high participants raised their foot during the motion, and there was no significant difference in the average step height between children and adults. The standard walk ratio for adults in the gait

literature is 0.0065 m/steps/min [112, 118]. However, we found a lower walk ratio for adults (M = 0.001 m/steps/min, SD = 0.001). The lower walk ratio is because the maximum step height that has been achieved in our study while moving in place is much less than the average step lengths noted in the literature (M = 0.68m) [118]. Children also had an average walk ratio of 0.001 m/steps/min (SD = 0.0009). We also found a significant effect of motion ($F_{3,57} = 50.47$, $p < 0.0001$). Post-hoc tests showed that the following motion pairs differed: walk/walk fast ($p < 0.001$). Participant's walk ratios were higher when walking in place compared to any of the other motions (see Table 5-2). This result follows from what we might expect: participants have more coordination and balance in the (slowest) walking motion when performing the motion in place.

5.2.5 Results–Temporal Features

We discuss the computation of time-based gait features, and present our findings using the same statistical analysis we used for the spatial features. Time-based features (except step speed) show significant effects by both age group and motion. Thus, these are promising features to use to differentiate between child and adult motion.

Step time (Figure 5-6A). The time for each step during an motion is computed as the difference between the time stamp for the end frame (when the foot is back on the ground) and the timestamp for the start frame (when the foot first leaves the ground). A two-way repeated measures ANOVA on step time with a between-subjects factor of age group (child, adult), and a within-subjects factor of motion (walk, walk fast, run, run fast) found a significant main effect of age group ($F_{1,18} = 12.15$, $p < 0.05$). Children move faster compared to adults (see Table 5-2). During our data collection, we observed that, given the same prompts, children were more energetic and enthusiastic when performing the motions compared to adults. We also found a significant effect of motion ($F_{3,54} = 55.86$, $p < 0.0001$). Post-hoc tests showed that the following motion pairs differed: walk/walk fast ($p < 0.001$), walk fast/run ($p < 0.05$). As expected, people have a faster step time when running, and become slower when walking fast, and walking, respectively (Table 5-2), validating our prompts. We also found a significant interaction effect ($F_{3,54} = 4.74$, $p < 0.05$). Children have a faster step time than adults when performing all the

motions except running fast (see Figure 5-6A). Since adults were less enthusiastic than children, the prompt "run as fast as you can" could have encouraged adults to finally "pick up the pace" and exert themselves more than they did in the previous motions.

Cycle time (Figure 5-6B). The time for a cycle is computed as the time it takes to complete two consecutive steps of the same foot. A two-way repeated measures ANOVA on cycle time with a between-subjects factor of age group (child, adult), and a within-subjects factor of motion (walk, walk fast, run, run fast) found a significant main effect of age group ($F_{1,18} = 7.61$, $p < 0.05$). Like step time, children complete cycles with a shorter time duration compared to adults (Table 5-2). We also found a significant effect of motion ($F_{3,57} = 102.13$, $p < 0.0001$). Post-hoc tests showed that the following motion pairs differed: walk/walk fast ($p < 0.001$), run/run fast ($p < 0.05$). Intuitively, people exhibit the fastest cycle time when running fast, and cycle time is slowest for walking (Table 5-2). Compared to step time, participants complete individual steps faster when running than when walking fast, but the time to complete successive steps (in this case two steps of the same foot) is roughly the same in both motions. Also, unlike step time, we found no significant interaction between age group and motion. This could be because cycle time includes motion of both feet, whereas our calculation of step time includes only the left foot. Previous research [113] has shown people exhibit a strength imbalance on their non-dominant side, which could lead to higher variability in the motion.

Cycle f (Figure 5-6C). The cycle frequency is computed as the inverse of the cycle time (1/cycle time). A two-way repeated measures ANOVA on cycle frequency with a between-subjects factor of age group (child, adult), and a within-subjects factor of motion (walk, walk fast, run, run fast) found a significant main effect of age group ($F_{1,18} = 10.53$, $p < 0.05$). Children have a higher cycle frequency compared to adults (Table 5-2). We also found a significant effect of motion ($F_{3,54} = 46.53$, $p < 0.0001$). Like cycle time, post-hoc tests showed that the following motion pairs differed: walk/walk fast ($p < 0.05$), run/run fast ($p < 0.05$). Unlike cycle time, we found a significant interaction effect ($F_{3,54} = 3.55$, $p < 0.05$) between age group and motion. Children have a higher cycle frequency than adults when performing all

motions, except walking in place (see Fig. 5-6c). Previous research has found lower cycle frequency is correlated to a lower physical energy cost [85], and young children are not as experienced at optimizing this cost as adults.

Step speed. The speed of a step is computed as the ratio of step height to step time. A two-way repeated measures ANOVA on step speed with a between-subjects factor of age group (child, adult) and a within-subjects factor of motion (walk, walk fast, run, run fast) found no significant main effect of age group ($F_{1,18} = 0.66$, *n.s.*), unlike all the other temporal features. We believe age group is not significant because the step speed is highly dependent on the step height, a distance-based feature ($r = 0.90$, $p < 0.0001$). However, we did find a significant effect of motion ($F_{3,57} = 19.06$, $p < 0.0001$). Post-hoc tests showed that the following motion pairs differed: run/run fast ($p < 0.001$). Step speed, being highly dependent upon step height, shows the same pattern by motion with respect to motion intensity.

Cadence (Figure 5-6D). The cadence is computed as the ratio of the total number of steps taken during a motion to the total time duration of that motion. A two-way repeated measures ANOVA on cadence with a between-subjects factor of age group (child, adult) and a within-subjects factor of motion (walk, walk fast, run, run fast) found a significant main effect of age group ($F_{1,18}= 5.86$, $p < 0.05$). The number of steps taken per minute for children is higher compared to adults (Table 5-2). This higher number is expected because we know from step time that children move faster than adults, and are more enthusiastic. Thus, it follows that children will complete more steps than adults in a similar time. Our analysis also show that adults have a lower cadence (M = 95 steps/min, SD = 19) when walking in place in our study. This value corresponds to a medium walking cadence (80 – 99 steps/min) asserted by Tudor-Locke et al. [130], but differs from the average cadence of 120 steps/min proposed by Gage et al. [39] for adults. This difference may be because adults did not walk long enough (<20 steps) in our study and may not have settled into a cadence. The finding could also be because participants in our study were walking in place compared to walking forward along a distance. Similarly, children in our study had an average cadence of (M = 103 steps/min, SD = 13) which is lower than the average cadence

of 138 steps/min in Dusing and Thorpe's study, for children ages 5 to 10 [35]. However, Barreira et al. [10] found that children spent more time in their open walking study walking at cadences of 100–119 steps/min compared to cadences of 120+ steps/min. We also found a significant effect of motion ($F_{3,54} = 55.75$, $p < 0.001$). Post-hoc tests showed that the following motion pairs differed: walk/walk fast ($p < 0.01$), run/run fast ($p < 0.05$). Cadence exhibits a similar pattern by motion as cycle frequency and cycle time: we discuss the relationship between these features in the next section. We also found a significant interaction effect ($F_{3,54} = 2.99$, $p < 0.05$). Children have a higher cadence than adults when performing all motions except walking (see Figure 5-6D). Like cycle frequency, children display the same pattern of cadence by motion, in which they are expending higher energy and exhibiting lower coordination than adults.

5.2.6 Discussion

Our results showed significant differences in temporal features such as step time, cycle time, cycle frequency, and cadence. Children have a faster step time and cycle time and a higher cadence, but a lower cycle frequency compared to adults. Hence, these features could be promising to use to differentiate between child and adult motion. We found no significant differences between child and adult motion for spatial features, showing that these features may help differentiate children's and adults' walking and running motions. Hence, children's motions are quantifiably different than adults' in ways that might affect recognition (e.g., speed and movement time), which may affect the recognition performance of children's motions, further emphasizing the need to tailor motion recognizers to children's motions. For example, the coordination of a child's movement depends on the motion being performed, as in our study the children had higher coordination evident during walking in place motions versus running fast motions. Recognizers will need higher tolerance for variance in the motion to be able to recognize less coordinated motion from children.

In addition to recognition, our findings also have implications for the design of whole-body motion applications for children. Our post-hoc analysis found no significant difference between walk fast and run in all the features except step time, and conversely, we found a significant

difference between run and run fast in all the features except step time. Furthermore, our results for cycle frequency and cadence show that children exhibit more energy than adults when performing the motions in our study. Taken together, we can assert that children exhibit higher exertions than adults for the same motion prompts. This finding suggests that designers of whole-body motion applications should tailor interaction prompts given to users based on the age group and desired level of exertion. For example, for higher levels of exertion, designers need only prompt children to "Run", but must prompt adults to "Run fast."

5.3 Summary

Findings from the studies presented in this chapter showed that the motion qualities that distinguish child motion from adult motion can be quantified. Although our first study showed that children are more inconsistent in how they perform motions compared to adults, the study did not provide any quantifiable features that can help inform a deeper understanding of the reasons why children are inconsistent in how they perform motions compared to adults. Even though our second study presents quantifiable features from the gait literature that differentiate child motion from adult motion, the features are not generalizable. Gait features are optimized for analyzing gait and rely on the periodicity of the motion, which makes them unsuitable for analyzing motions that are not periodic, such as exercise motions (e.g., "a Jump" or "Kick"). In the next chapter, we focus on identifying a set of features that can be applied to a broader set of motions. Using these features, we aim to quantify the differences between children's and adults' motions to establish a set of features that characterize children's natural motion qualities. Through this understanding, we aim to propose design guidelines for tailoring motion recognition systems to children's motion qualities to enable accurate recognition of their motions as well as propose guidelines for the design of whole-body gesture sets and whole-body gesture applications for children.

CHAPTER 6
CHARACTERIZING CHILDREN'S NATURAL MOTION QUALITIES

So far, we have established that children's motions differ from adults' motions. In this chapter, we continue our goal of understanding how children move differently from adults (i.e., their natural motion qualities). We do this by identifying a set of features that quantitatively describe motions and evaluate them on a subset of children's and adults' motion to reveal differences. We also qualitatively analyzed children's motions to provide support for our quantitative results.

6.1 Quantifying Differences Between Child and Adult Motion

To identify features that are applicable to motions regardless of the motion type, we propose a set of articulation features that can quantitatively describe motions. Specifically, we use human-readable features proposed by Vatavu [132] for characterizing human motion performance, which we refer to as "global-level features". However, because global-level features focus on the whole-body, they will not be as helpful characterizing properties of individual joints that make up the motion. Hence, we also propose a new set of features that characterize geometric properties such as length, shape, and curvature, of motion paths of a joint, which we termed "joint-level features".

6.1.1 Global-Level Features

Vatavu [132] identified 17 features for characterizing human motion performance comprising spatial features (dependent on distance), kinematic features (dependent on time) and appearance features (dependent on the composition of postures). We refer to these features as "global-level features" because they describe motions globally, based on the overall posture or pose of the body (i.e., the positioning of the body at a specific time instance as defined by a set of joints with positions in 3D space, see Figure 3-1B in Chapter 1).

6.1.1.1 Spatial features

These features capture motion qualities related to the area, volume, and amplitude of gesture movement performed by the whole-body or body parts [132]. The main features included in this category are "gesture volume", "gesture area", and "quantity of movement" (i.e., how much movement the user performs), "difference of movement" and "ratio of movement". Vatavu

[132] also derived several features from these main features, which the author notes depends on what the research is aiming to investigate. For example, Vatavu [132] derived the "quantity of hands movement", "difference of hands movement", and "ratio of hands movement from the quantity of movement, difference of movement, and ratio of movement respectively. The author notes that these features are of interest to a researcher investigating the relationship between the dominant and non-dominant hand during movement. Vatavu [132] also derived the "ratio of hands to legs movement" and "ratio of hands to body movement" for researchers who are interested in the relationship between the hands and legs and hands and body, respectively. Because some motions in the Kinder-Gator dataset [3] are heavily dependent on the upper body (e.g., "Raise your hand") while others are heavily dependent on the lower body (e.g., "Bend your knee"), we are more interested in how the upper body moves in relation to the lower body rather than how one limb (e.g., hands) moves in relation to another limb (e.g., legs). Therefore, in addition to the main features above, we derived the following features: "quantity of upper body movement" and "quantity of lower body movement" to capture the relationship between the upper and lower body during movement. We describe the main features and derived features below:

Gesture volume (GV). Vatavu defines this feature as the volume of the 3–D space in which the motion is performed [132]. It is computed as the product of the difference between the maximum and minimum positions of the body in the x (length), y (height), and z (depth) dimensions:

$$GV = \prod_{\delta \in \{x,y,z\}} (max_{i,j}\{\delta_j^i\} - min_{i,j}\{\delta_j^i\}) \tag{6-1}$$

Where δ represents each of the dimensions along which the motion is tracked (i.e., x, y, z), i enumerates all body poses of the motion, and j enumerates all 20 joints tracked by the Kinect.

Gesture area (GA). Vatavu defines this feature as the area of the 2–D space in front of the motion sensor where the motion is performed [132]. It is computed as the product of the difference between the maximum and minimum positions of the body in the x (length) and y

(height) dimensions only:

$$GV = \prod_{\delta \in \{x,y\}} (max_{i,j}\{\delta_j^i\} - min_{i,j}\{\delta_j^i\}) \tag{6-2}$$

Where δ represents each of the x and y dimensions, i enumerates all body poses of the motion, and j enumerates all 20 joints tracked by the Kinect.

Quantity of movement ($\mathbf{Q}_M$). Vatavu defines this feature as the total amount of movement performed by the user [132]. It is computed as the cumulative pairwise Euclidean distance between corresponding joints of time–consecutive frames of the data:

$$Q_M = \frac{1}{\lambda} \sum_{i=2}^{n} \sum_{j=1}^{J} ||p_j^i - p_j^{i-1}|| \tag{6-3}$$

Where $||p_j^i - p_j^{i-1}||$ is the Euclidean distance between a joint j of a body pose p^i in the current time frame i and the same joint of a body pose p^{i-1} in the previous time frame $i-1$ defined as two 3D points [132]. λ is a normalization factor such that if λ =1, then the result is the cumulative quantity of movement and if λ =number of tracked joints, then the result is the average quantity of movement per joint.

The generalized version of this feature, also known as the "Generalized Quantity of Movement", weights the Euclidean distance so that certain joints can be emphasized over others [132]:

$$Q_M = \frac{1}{\lambda} \sum_{i=2}^{n} \sum_{j=1}^{J} (w_j).||p_j^i - p_j^{i-1}|| \tag{6-4}$$

Where w_j is any value representing the weight of a joint j. From this feature, we derived the following two features:

Quantity of upper body movement ($\mathbf{Q}_{MU}$). This feature quantifies the amount of movement in the upper body. We calculated this feature using the same formula as the generalized quantity of movement with w_j =1 for joints in the upper body, namely: head, shoulder center, left shoulder, right shoulder, left elbow, right elbow, left wrist, right wrist, left hand, right hand, and w_j =0 for joints in the lower body.

Quantity of lower body movement ($\mathbf{Q}_{ML}$). This feature quantifies the amount of movement in the lower body. Similar to the upper body movement feature, we calculated this feature using the same formula as the generalized quantity of movement with $w_j = 1$ for joints in the lower body, namely: left hip, right hip, left knee, right knee, left ankle, right ankle, left foot, and right foot, and $w_j = 0$ for joints in the upper body.

Difference of movement ($\mathbf{D}_M$). Vatavu defines this feature as the difference between the quantity of movement of two different body parts [132]:

$$D_M = Q_M(\text{first body part}) - Q_M(\text{second body part}) \tag{6-5}$$

As mentioned earlier, our main interests are in the upper and lower body movement, so I am deriving the following feature:

Difference of body movement ($\mathbf{D}_{body}$). We calculated this feature as the difference between the quantity of movement in the upper body and lower body:

$$D_{body} = Q_{MU} - Q_{ML} \tag{6-6}$$

Ratio of movement ($\mathbf{R}_M$). Vatavu defines this feature as the ratio of the quantity of movement in one body part with respect to the quantity of movement in another body part [132]:

$$R_M = \frac{Q_M(\text{first body part})}{Q_M(\text{second body part})} \tag{6-7}$$

Ratio of body movement ($\mathbf{R}_{body}$). We derived this feature from R_M. We calculated this feature as the ratio of the quantity of movement in the upper body to the quantity of movement in the lower body:

$$R_M = \frac{Q_{MU}}{Q_{ML}} \tag{6-8}$$

6.1.1.2 Kinematic features

This set of Vatavu's features capture motion qualities related to the time it takes to perform a motion and the speed at which the motion is performed [132]. Kinematic features only consider the duration when the motion of interest is performed. Therefore, we exclude periods when

addition motions that are not of interest are being performed. For example, during the collection of the Kinder-Gator dataset [3], participants were first required to complete a T-pose before performing the actual motion. Hence, for the computation of the Kinematic features, the T-pose will be excluded from the motion. Vatavu [132] identified two kinematic features, namely "Performance Time (T)" and "Average Gesture Speed (S)":

Performance time. Vatavu defines this feature as the time it takes the user to perform the motion [132]. It is computed as the difference between when the user starts the motion and when the user ends the motion, reported in seconds:

$$T = T_N - T_1 \tag{6-9}$$

Where T_N is the timestamp of the last frame and T_1 is the timestamp of the first frame. Recall that a frame is the position of all joints tracked by the motion sensor in 3D space at a given time instance.

Average gesture speed (S). Vatavu defines this feature as a measure of how fast or slow the user is when performing the motion [132]. It is computed as the ratio of the quantity of movement (a spatial feature) and the performance time:

$$S = \frac{Q_M}{T} \tag{6-10}$$

6.1.1.3 Appearance features

This set of Vatavu's features characterize how motions decompose into simple units of movements [132]. They include "Body Posture Variation (BPV)", "Body Posture Diffusion (BPD)", "Body Posture Density (BPρ)", and "Body Posture Rate (BPR)".

Body posture variation (BPV). Vatavu defines this feature as the average deviation of a body posture from the centroid posture of the motion [132]. A posture, also known as a pose or a frame, is the position of all the joints tracked by the motion sensor in 3D space at a given time

instance. The centroid pose is calculated as the average position of each in 3D space:

$$BPV = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{J}||p_j^i - \overline{p_j}|| \tag{6-11}$$

Where $||p_j^i - \overline{p_j}||$ is the Euclidean distance between each pose of the motion and n is the number of poses.

Body posture diffusion (BPD). Vatavu defines this feature as the maximum difference between the body poses that make up the motion [132]. It is computed by finding the difference between each pose and all corresponding poses and selecting the maximum difference:

$$BPD = max_{1\leq<k\leq n}\left\{\sum_{j=1}^{J}||p_j^i - p_j^k||\right\} \tag{6-12}$$

Where p_j^i and p_j^k are corresponding joints of pose p^i at time frame i and pose p^k at time frame k respectively.

Body posture density (BPρ). Vatavu defines this feature as the variation of the body pose over the 3D space in which the motion is performed [132]. It is computed as the ratio of the Body Pose Variation to the Gesture Volume (a spatial feature):

$$BP\rho = \frac{BPV}{GV} \tag{6-13}$$

Body posture rate (BPR). Vatavu defines this feature as the variation of the body pose over the time it took to produce the motion [132]. It is computed as the ratio of the Body Pose Variation to the Performance Time (a kinematic feature):

$$BPR = \frac{BPV}{T} \tag{6-14}$$

In total, our work presents 13 global-level features from prior work [132] that quantitatively describes users' motion performance and which we use to analyze child and adult motions, namely: Gesture Volume, Gesture Area, Quantity of Movement, Quantity of Upper Body Movement, Quantity of Lower Body Movement, Difference of Movement, Ratio of Movement, Performance Time, Average Gesture Speed, Body Pose Variation, Body Pose Diffusion, Body

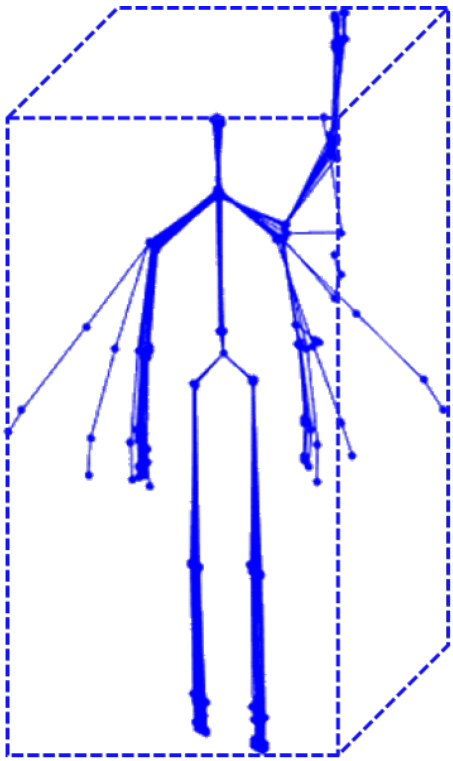


Figure 6-1. Gesture Volume computed on postures of a Raise your hand motion

Pose Density, and Body Pose Rate. Fig. 6-1 shows an example of a global-level feature computation on the postures of a motion.

6.1.2 Joint-Level Features

The limitation of the global-level features described above is that they focus on the position of the whole-body at a given point in time, so these features will not characterize properties of individual joints that are critical to performing motions. For example, consider a user performing the "Jump" motion. Global-level features will be useful in ensuring that the user is performing the "Jump" motion as opposed to a different motion (e.g., "Do a forward lunge" motion). However, global-level features might not be as helpful in understanding whether two users moved individual joints responsible for making movements, such as limb movements [92], differently as they perform the motion. For example, users are expected to lift both feet off the ground to perform the "Jump" motion. Depending on how users control their feet, there could be variations in how users lift their feet off the ground, which could result in differences in how these users

perform the "Jump" motion. In this example, global-level features may not be as helpful in understanding whether two users lifted their feet differently when performing the "Jump" motion. Therefore, global-level features will not capture motion qualities that relate to subtleties of the joint articulation path (i.e., as joint-level features can); the joint articulation path is defined by a set of consecutive 3D points the joint moves through along the time-domain. These subtleties can inform an understanding of the variations in how different users move their joints during motions. For this reason, in addition to the motions proposed by Vatavu [132], we identified a set of joint-level features that quantify geometric properties of the joint articulation paths necessary for performing the motions, such as length, shape, and curvature. We referred to these features as "geometric features". These features were inspired by the relative accuracy features from Vatavu et al. [135], features for tracking mouse paths [56], and features from Laban Movement Analysis [72]. To identify the joints necessary to perform motions, we use our filterJoint method discussed in Chapter 5, which uses standard deviation and K-means clustering [53] iteratively to select the set of joints that are actively moving during a motion. In this section, we describe a method for identifying geometric features and the geometric features that emerged from this method.

6.1.2.1 Joint task axis

To identify features that can describe geometric properties of joint articulation paths, we rely on a method from a closely related field of research, 2D stroke gesture recognition. Vatavu et al. [135] proposed a method that can be used to measure the inconsistency between stroke gesture articulation paths, called the "gesture task axis". The authors defined the gesture task axis as a representative way to articulate a stroke gesture [135] and proposed three types of gesture task axis:

1. Geometric gesture task axis, which is defined by the designer using geometric primitives such as lines and curves [135].
2. Average gesture task axis, which the authors define as the "average shape of a set of user-captured gesture samples" [135].
3. Template gesture task axis, which the authors define as a "canonical template form supplied to a recognizer to which articulated gestures will be compared in a template-based

matching approach" [135]. Given a dataset of stroke gestures (e.g., for the letter A), the template task axis is the gesture with the least distance to a representative gesture, computed as the average of all the gestures in the dataset [135].

Vatavu et al. [134] further noted that the stroke gesture paths need to be resampled to the same number of points to enable point-to-point comparison and translated so that the centroid is at the origin, before computing a 2D stroke gesture task axis,.

Similarly, we defined a "joint task axis" as a representative way to move a joint. Unlike stroke gestures in which there is only one articulation path, defined by the finger's movement, and one task axis, motions have multiple articulation paths, defined by all the joints tracked by the motion sensor. Therefore, a motion will have multiple joint task axes, one for each joint necessary to perform the motion. To define the joint task axis, we used the template gesture task axis method. The geometric gesture task axis requires designers to defined the representative path. However, unlike stroke gestures wherein the lines and curves that make up the gesture (e.g., the letter "A") are already defined and well-known by designers, the lines and curves that make up a joint's articulation path are not as well-defined. Therefore, it will be difficult for designers to intuit the expected path of joint (e.g., the knee joint) in 3D space as a user moves that joint, which means that the joint path resulting from a geometric gesture tasks axis may not be representative of how that joint moves. Similarly, we do not use the average gesture task axis method because the average of all the joint paths of a joint does not guarantee a path that a user will actually move that joint through during a motion. Consequently, the combination of the average joint paths for all joints may not maintain the integrity of the motion (e.g., relative positioning between body parts). Therefore, results from features that depend on the geometric gesture task axis or average gesture task axis could be inaccurate in quantitatively describing a user's motion.

To apply the template method, we use a leave-one-out approach inspired by the Leave-One-Out-Cross-Validation approach (LOOCV) used in recognition experiments [4]. In LOOCV, motions from one participant are selected for testing (i.e., the candidate user) while motions from all other participants are used for training. This process is repeated until all participants have been selected once for testing. Similarly, given a set of motions (e.g., for the

motion Raise your hand), one user is selected as the candidate user. For each of their joints that is actively moving as selected by the filterJoint method [5], the template joint task axis for that joint will include every other user's corresponding joint articulation path in the motion set. For example, given three users (c1, c2, c3) wherein c1 is the candidate user who actively moves their right hand when raising their hand, then the task axis for the right hand joint will include the articulation paths of the right hand joint for users c2 and c3. The process is repeated until every user in the set has been selected as the candidate user once and all the joint task axes for all the joints that each candidate user is actively moving has been generated. We formalize the joint task axis as follows:

a. Given a set of motions of a specific type
$M = \{m_1, m_2, m_3, ...m_U | U = \text{number of users that performed } T\}$, where each motion in M is represented by J joints that are actively moving as selected by the filterJoint method. Then for a specific candidate user $C \in U$ with motion m_i in M with q joints that are actively moving:
$$J_C = \{j_C^1, j_C^2, j_C^3, ...j_C^q\} \tag{6-15}$$
Where each joint j_C^k is defined by a series of n 3D points
$$p = \{p_1, p_2, p_3, ..., p_n | p_i = (x_i y_i, z_i)\} \tag{6-16}$$

b. The joint task axis T_σ for j_C^k will include the corresponding joint articulation paths of every other user in U:
$$T_\sigma^{j_C^k} = \left\{ j_1^k, j_2^k, j_3^k, ..j_{U-1}^k | j_i^k \neq j_C^k \right\} \tag{6-17}$$

c. Then the joint task axis T_σ for J_C is: T
$$T_\sigma = \left\{ T_\sigma^{j_C^1}, T_\sigma^{j_C^2}, T_\sigma^{j_C^3}, ..., T_\sigma^{j_C^q} \right\} \tag{6-18}$$

6.1.2.2 Geometric features

Next, we define a set of joint-level features that characterize the deviation of the articulation path of a joint from the joint task axis with respect to properties, such as length, shape, and curvature. These features were inspired by the relative accuracy features from Vatavu et al. [135], features for tracking mouse paths [56], and features from Laban Movement Analysis [72, 155]. Eight of the eleven features we identified rely on a concept of "error", which does not imply that

the user moved the joint in the wrong way, but rather measures inconsistency with respect to the task axis [135]. Each feature requires a comparison between the joint path of the candidate user and every other users' corresponding joint path (i.e., the articulation paths in the joint task axis). Then, the average of the feature computation across all the comparisons and all the joints that the candidate user is actively moving is used to compute the feature for a given candidate user's motion instance. For example, given three users (c1, c2, c3) where c1 is the candidate user who actively raises their right hand and right elbow when raising their hand. To compute joint-level feature f, we compare c1's right hand path to the right hand path of c2 and c1's right hand path to the right hand path of c3 and compute the average (u1). We do the same for the right elbow joint to compute the average (u2). Then, we take the average of u1 and u2 as the value of feature f.

To ensure accurate comparison of a joint articulation path of a candidate user and a joint in the joint task axis belonging to a representative user, we apply the following methods to address the four ways in which motions can be performed that will impact the distance between two joint articulation paths:

a. Same joint (i.e., left vs right) and same direction, compare as-is. For example, both candidate user and representative user swipe their right hand from left to right.

b. Same joint but different directions (e.g., swiping right to left vs. swiping left to right), FLIP the joint articulation path 180^o along the x-axis to change direction and then compare joint articulation path and joint task axis. For example, both candidate user and representative user swipe their right hand, but the candidate user swipes from right to left while the representative user swipes from left to right.

c. Different joints but same direction, REPLACE articulation path of the joint task axis with the articulation path of the joint of its opposite limb (i.e., replace left joints with right joints and vice-versa, leave middle joints as-is) then compare. For example, both candidate and representative user swipe from right to left, but the candidate user swipes with their right hand while the representative user swipes with their left hand.

d. Different limbs and different directions, FLIP the joint articulation path, REPLACE articulation path of joint task axis, then compare. For example, the candidate user swipes with their left hand from right to left while the representative user swipes with their right hand from left to right.

For a given candidate joint articulation path and an articulation path in the set of articulation paths in the corresponding joint task axis, the comparison with the least Euclidean distance is used to compute all the joint-level features for the pair of joint articulation paths being compared.

Shape error. This feature measures the average absolute deviation of the shape of a user's joint path in a motion instance from the shape of the same joint path in the joint task axis. It is measured as the average Euclidean distance between each 3D point in the given joint and each 3D point in the joint task axis. Given a motion instance for candidate user C,an actively moving joint j_C^k, and a set of joint articulation paths for the corresponding joint task axis $T_\sigma^{j_C^k}$, then the shape error is computed as: [5].

$$\begin{aligned} SHE(C) &= \frac{1}{q}\sum_{k=1}^{q} SHE(j_C^k) \\ SHE(j_C^k) &= \frac{1}{U-1}\sum_{u=1}^{U-1} SHE(j_C^k, j_u^k) \forall j_i^k \in T_\sigma^{j_C^k} \end{aligned} \tag{6-19}$$

$$SHE(j_C^k, j_u^k) = \frac{1}{n}\sum_{i=1}^{n} min\Bigg(||j_{c(i)}^k - j_{u(i)}^k||, ||flip(j_{c(i)}^k) - j_{u(i)}^k|| \\ , ||flip(j_{c(i)}^k) - mirror(j_{u(i)}^k)||, ||j_{c(i)}^k - mirror(j_{u(i)}^k)|| \Bigg)$$

Where $SHE(j_C^k, j_u^k)$ is the shape error between two joint articulation paths, $SHE(j_C^k)$ is the shape error over all the joint articulation paths in the joint task axis, and $SHE(C)$ is the shape error over all the actively moving joints. q is the number of joints that are actively moving for an instance of the motion from C, and U is the number of users in the dataset. Similar to the 2D stroke gesture task axis, each joint articulation path is first resampled to the same number of n points and translated so that the centroid is at the origin to enable point-to-point correspondence.

Shape variability. This feature measures how uniform the shape errors are along a joint articulation path and is computed as the standard deviation of the distances between each 3D point in the joint path in a motion instance and each 3D point of a joint path in the joint tasks axis. Given a motion instance for candidate user C,an actively moving joint j_c^k, and a set of joint articulation paths for the corresponding joint task axis $T_\sigma^{j_C^k}$, then the shape variability is computed

as:

$$
\begin{aligned}
SHV(C) &= \frac{1}{q}\sum_{k=1}^{q} SHV(j_C^k) \\
SHV(j_C^k) &= \frac{1}{U-1}\sum_{u=1}^{U-1} SHV(j_c^k, j_u^k) \forall j_i^k \in T_\sigma^{j_C^k} \\
SHV(j_C^k, j_T^k) &= \frac{1}{n-1}\sqrt{\sum_{i=1}^{n} ||j_{c(i)}^k - j_{u(i)}^k|| - SHE(j_C^k, j_u^k)}
\end{aligned}
\tag{6-20}
$$

Where $SHV(j_C^k, j_u^k)$ is the shape error between two joint articulation paths, $SHV(j_C^k)$ is the shape error over all the joint articulation paths in the joint task axis, and $SHV(C)$ is the shape error over all the actively moving joints.

Bend error. Vatavu et al. [135] defined this feature as a user's tendency to bend strokes during the articulation of a stroke gesture. With respect to motions, we define this feature as a user's tendency to bend or curve the articulation paths of their joints when performing the motion. It is computed as the average of the absolute difference between the turning angle of a user's joint path in a motion instance and the turning angle of a joint path in the joint task axis. The turning angle at a 3D point p is the angle, θ, between p and the previous point $(p-1)$ and p and the next point $(p+1)$ [135]. Given a motion instance for candidate user C, an actively moving joint j_C^k, and a set of joint articulation paths for the corresponding joint task axis $T_\sigma^{j_C^k}$, then the bend error is computed as:

$$
\begin{aligned}
BE(C) &= \frac{1}{q}\sum_{k=1}^{q} BE(j_C^k) \\
BE(j_C^k) &= \frac{1}{U-1}\sum_{u=1}^{U-1} BE(j_C^k, j_u^k) \forall T \in T_\sigma^{j_C^k} \\
BE(j_C^k, j_u^k) &= \frac{1}{n}\sum_{i=1}^{n} |\theta_{c(i)}^k - \theta_{u(i)}^k|
\end{aligned}
\tag{6-21}
$$

Where $BE(j_C^k, j_u^k)$ is the bend error between two joint articulation paths, $BE(j_C^k)$ is the bend error over all the joint articulation paths in the joint task axis, and $BE(C)$ is the bend error over all the

actively moving joints and $\theta^k_{c(i)}$ and $\theta^k_{u(i)}$ are the turning angles computed as:

$$\theta_p =< (\overline{p-1.p}, \overline{p.p+1}) \tag{6-22}$$

Bend variability. This feature measures the uniformity of the bending errors along a joint path and is computed as the standard deviation of the differences between the turning angles of a joint path in a motion instance and the turning angle of a joint path in the joint task axis. Given a motion instance for candidate user C, an actively moving joint j^k_C, and a set of joint articulation paths for the corresponding joint task axis $T_\sigma^{j^k_C}$, then the bend variability is computed as:

$$\begin{aligned} BV(C) &= \frac{1}{q}\sum_{k=1}^{q} BV(j^k_C) \\ BV(j^k_C) &= \frac{1}{U-1}\sum_{u=1}^{U-1} BV(j^k_C, j^k_u) \forall T \in T_\sigma^{j^k_C} \\ BV(j^k_C, j^k_u) &= \frac{1}{n-1}\sqrt{\sum_{i=1}^{n} |\theta^k_{c(i)} - \theta^k_{u(i)}| - BE(j^k_C, j^k_u)} \end{aligned} \tag{6-23}$$

Where $BV(j^k_C, j^k_u)$ is the bend error between two joint articulation paths, $BV(j^k_C)$ is the bend error over all the joint articulation paths in the joint task axis, and $BV(C)$ is the bend error over all the actively moving joints.

Length error: Vatavu et al. [134] defined this feature as a user's tendency to stretch gesture strokes with respect to the task axis. For motions, we defined this feature as a user's tendency to stretch their joint articulation paths. It is computed as the absolute difference between the path lengths of a joint path in a motion instance and a joint path in the joint task axis. Given a motion instance for candidate user C, an actively moving joint j^k_C, and a set of joint articulation paths for the corresponding joint task axis $T_\sigma^{j^k_C}$, then the length error is computed as:

$$\begin{aligned} LE(C) &= \frac{1}{q}\sum_{k=1}^{q} LE(j^k_C) \\ LE(j^k_C) &= \frac{1}{U-1}\sum_{u=1}^{U-1} LE(j^k_C, j^k_u) \forall T \in T_\sigma^{j^k_C} \\ LE(j^k_C, j^k_u) &= |L(j^k_C) - L(j^k_u)| \end{aligned} \tag{6-24}$$

Where $LE(j_C^k, j_u^k)$ is the length error between two joint articulation paths, $LE(j_C^k)$ is the length error over all the joint articulation paths in the joint task axis, and $LE(C)$ is the length error over all the actively moving joints and $L(j_C^k)$ and $L(j_u^k)$ are the path lengths of the joint articulation paths being compared.

Size error. Vatavu et al. [135] defined this feature as users' tendency to stretch gesture strokes with respect to the gesture area size, which is the area of the smallest bounding box that can encompass the gesture path. Because motions occur in 3D space, we define this feature as users' tendency to stretch the articulation paths of joints with respect to joint volume size. It is computed as the absolute difference between the volume of the smallest bounding box encompassing the joint path of a motion instance and a joint path in the task axis. Given a motion instance for candidate user C, an actively moving joint j_C^k, and a set of joint articulation paths for the corresponding joint task axis $T_\sigma^{j_C^k}$, then the size error is computed as:

$$
\begin{aligned}
SE(C) &= \frac{1}{q}\sum_{k=1}^{q} SE(j_C^k) \\
SE(j_C^k) &= \frac{1}{U-1}\sum_{u=1}^{U-1} SE(j_C^k, j_u^k) \forall T \in T_\sigma^{j_C^k} \\
SE(j_C^k, j_u^k) &= |V(j_C^k) - V(j_u^k)|
\end{aligned}
\tag{6-25}
$$

Where $SE(j_C^k, j_u^k)$ is the size error between two joint articulation paths, $SE(j_C^k)$ is the size error over all the joint articulation paths in the joint task axis, and $SE(C)$ is the size error over all the actively moving joints and $V(j_C^k)$ and $V(j_u^k)$ are the volumes of the bounding boxes of the joint articulation paths.

Efficiency. This feature measures how efficient a path is in achieving its goal [56]. In mouse tracking, this feature is computed as the ratio of the length of the shortest path (defined as a straight line) to the length of the curve. With respect to motions, we defined efficiency as the ratio of the length of a joint path in the joint task axis, which is the ideal path to the length of a joint path in a motion instance. Given a motion instance for candidate user C, an actively moving joint j_C^k, and a set of joint articulation paths for the corresponding joint task axis $T_\sigma^{j_C^k}$, then the

efficiency is computed as:

$$
\begin{aligned}
E(C) &= \frac{1}{q}\sum_{k=1}^{q} E(j_C^k) \\
E(j_C^k) &= \frac{1}{U-1}\sum_{u=1}^{U-1} E(j_C^k, j_u^k) \forall T \in T_\sigma^{j_C^k} \\
E(j_C^k, j_u^k) &= \frac{L(j_u^k)}{L(j_C^k)}
\end{aligned}
\tag{6-26}
$$

Where $E(j_C^k, j_u^k)$ is the efficiency between two joint articulation paths, $E(j_C^k)$ is the efficiency over all the joint articulation paths in the joint task axis, and $E(C)$ is the efficiency over all the actively moving joints.

Time error: According to Vatavu et al. [135], this feature measures the difference between the total time it takes to articulate a stroke gesture and the task axis. We computed this feature as the difference between the total time it takes to articulate a joint path in a motion instance and the corresponding joint path in a joint task axis. Given a motion instance for candidate user C, an actively moving joint j_C^k, and a set of joint articulation paths for the corresponding joint task axis $T_\sigma^{j_C^k}$, then the time error is computed as:

$$
\begin{aligned}
TE(C) &= \frac{1}{p}\sum_{k=1}^{p} TE(j_C^k) \\
TE(j_C^k) &= \frac{1}{U-1}\sum_{u=1}^{U-1} TE(j_C^k, j_u^k) \forall T \in T_\sigma^{j_C^k} \\
TE(j_C^k, j_u^k) &= |time(C) - time(u)| \\
time(j^k) &= t_n - t_1
\end{aligned}
\tag{6-27}
$$

Where $TE(j_C^k, j_u^k)$ is the time error between two joint articulation paths, $TE(j_C^k)$ is the time error over all the joint articulation paths in the joint task axis, and $TE(C)$ is the time error over all the actively moving joints and t_n and t_1 are the timestamps of the last frame and first frame of a joint path j^k respectively. Since all the joints are tracked concurrently, the total time it takes to move one joint in the motion will be same as the total it takes to move all other joints, which

means that it is the time it takes the user to perform a motion (i.e., performance time, see section on global-level features).

Speed error. This feature measures the average difference between the speed of a joint path in a motion instance and the speed of a joint path in the joint task axis, where speed is the ratio of the quantity of movement, computed using the path length, to the time it takes to perform the motion (performance time). Given a motion instance for candidate user C, an actively moving joint j_C^k, and a set of joint articulation paths for the corresponding joint task axis $T_\sigma^{j_C^k}$, then the speed error is computed as:

$$\begin{aligned} VE(C) &= \frac{1}{q}\sum_{k=1}^{q} VE(j_C^k) \\ VE(j_C^k) &= \frac{1}{U-1}\sum_{u=1}^{U-1} VE(j_C^k, j_u^k) \forall T \in T_\sigma^{j_C^k} \\ VE(j_C^k, j_u^k) &= \frac{1}{n}\sum_{i=1}^{n} |S(j_C^k, i) - S(j_u^k, i)| \end{aligned} \tag{6-28}$$

Where $VE(j_C^k, j_u^k)$ is the speed error between two joint articulation paths, $VE(j_C^k)$ is the speed error over all the joint articulation paths in the joint task axis, and $VE(C)$ is the speed error over all the actively moving joints and $S_{C(i)}^k$ - $S_{u(i)}^k$ are the speeds of the joint articulation paths, computed as the ratio of the length of the joint path at frame i (also known as arc length) to the time duration at i. From Laban Movement Analysis [72], the speed of a joint j^k at frame i can be computed as:

$$V(j,i) = \begin{cases} \frac{arc-length(j_{i+1}^k) - arc-length(j_i^k)}{t_{i+1}-t_i} & i = 1 \\ \frac{arc-length(j_i^k) - arc-length(j_{i-1}^k)}{t_i - t_{i-1}} & i = n \\ \frac{arc-length(j_{i+1}^k) - arc-length(j_{i-1}^k)}{t_{i+1}-t_{i-1}} & \text{otherwise} \end{cases} \tag{6-29}$$

$$arc-length(j_i^k) = \sum_{a=2}^{i} ||j_a^k - j_{a-1}^k||$$

Where $i, a \in frames\ 1...n$ of joint j^k.

Acceleration error. This feature measures the difference in the acceleration of a joint path in a motion instance and the acceleration of a joint path in the corresponding task axis for that joint, where acceleration is the ratio of speed to performance time. Given a motion instance for candidate user C, an actively moving joint j_C^k, and a set of joint articulation paths for the corresponding joint task axis $T_\sigma^{j_C^k}$, then the acceleration error is computed as:

$$\begin{aligned} AE(C) &= \frac{1}{q}\sum_{k=1}^{q} AE(j_C^k) \\ AE(j_C^k) &= \frac{1}{U-1}\sum_{u=1}^{U-1} AE(j_C^k, j_u^k) \forall T \in T_\sigma^{j_C^k} \\ AE(j_C^k, j_u^k) &= \frac{1}{n}\sum_{i=1}^{n} |A(j_C^k, i) - A(j_u^k, i)| \end{aligned} \tag{6-30}$$

Where $AE(j_C^k, j_u^k)$ is the acceleration error between two joint articulation paths, $AE(j_C^k)$ is the acceleration error over all the joint articulation paths in the joint task axis, and $AE(C)$ is the acceleration error over all the actively moving joints and $A_{c(i)}^k$ - $A_{T(i)}^k$ are the accelerations of the joint paths, computed as the ratio of the velocity of the joint at i and the time duration at i. From Laban Movement Analysis [72], the acceleration of a joint j^k at frame i can be computed as:

$$A(j,i) = \begin{cases} \frac{S(j,i+1)-S(j,i)}{t_{i+1}-t_i} & i = 1 \\ \frac{S(j,i)-S(j,i-1)}{t_i-t_{i-1}} & i = n \\ \frac{S(j,i+1)-S(j,i-1)}{t_{i+1}-t_{i-1}} & \text{otherwise} \end{cases} \tag{6-31}$$

Jerk error. This feature measures the difference in the jerk of a joint path and the jerk of a joint path in the corresponding task axis for that joint, where jerk is the ratio of acceleration to performance time. Given a motion instance for candidate user C, an actively moving joint j_C^k, and a set of joint articulation paths for the corresponding joint task axis $T_\sigma^{j_C^k}$, then the jerk error is

computed as:

$$
\begin{aligned}
JE(C) &= \frac{1}{p}\sum_{k=1}^{p} JE(j_C^k) \\
JE(j_C^k) &= \frac{1}{U-1}\sum_{u=1}^{U-1} JE(j_C^k, j_u^k) \forall T \in T_{\sigma_C^{j^k}} \\
JE(j_C^k, j_u^k) &= \frac{1}{n}\sum_{i=1}^{n} |J(j_C^k, i) - J(j_u^k, i)|
\end{aligned}
\tag{6-32}
$$

Where $JE(j_C^k, j_u^k)$ is the jerk error between two joint articulation paths, $JE(j_C^k)$ is the jerk error over all the joint articulation paths in the joint task axis, and $JE(C)$ is the jerk error over all the actively moving joints and $J_{c(i)}^k$ - $J_{T(i)}^k$ are the jerks of the joint articulation paths, computed as the ratio of the acceleration of the joint at i to the time duration at i. From Laban Movement Analysis [72], the jerk of a joint j^k at frame i can be computed as:

$$
J(j,i) = \begin{cases}
\frac{A(j,i+1)-A(j,i)}{t_{i+1}-t_i} & i = 1 \\
\frac{A(j,i)-A(j,i-1)}{t_i-t_{i-1}} & i = n \\
\frac{A(j,i+1)-A(j,i-1)}{t_{i+1}-t_{i-1}} & \text{otherwise}
\end{cases}
\tag{6-33}
$$

In total, we proposed 11 joint-level features that quantitatively describe properties of a joint articulation path: Shape error, Shape Variability, Bend Error, Bend Variability, Length Error, Size Error, Efficiency, Time Error, Speed Error, Acceleration Error, and Jerk Error, which we then combine with the global-level features from prior work [132] to analyze children's and adults' motions. Fig. 6-2 shows an example of a joint-level feature computation on a joint articulation path of a motion.

6.1.3 Analysis

Our analysis aims to identify the features that differentiate child motion from adult motion. We computed both the global- and joint-level features on the same subset of motions from the Kinder-Gator dataset presented in Table 4-1 from Chapter 4. Recall that the motions in this subset include:

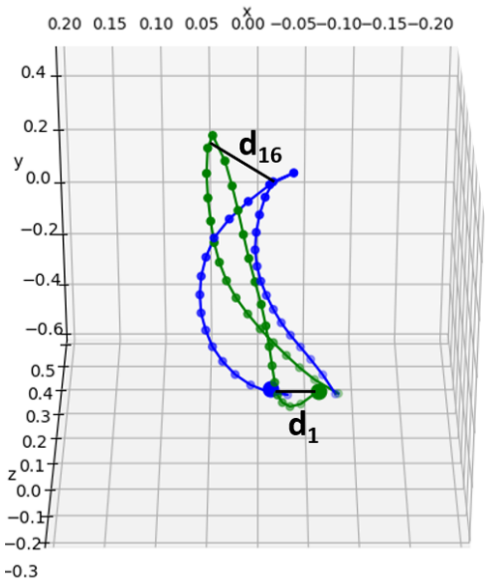


Figure 6-2. Shape Error (sum of d's) computed by comparing two articulation paths of the wrist joint of a Raise your hand motion where the blue path is the joint task axis on postures of a Raise your hand motion

1. **Upper-Body Motions**: "Raise your hand (RYH)", "Point at the camera (PAC)", "Raise your arm to one side (RAS)", "Put your hands on hips and lean to one side (PHL)", "Punch (P)", "Throw a ball as far as you can (TBF)", "Swipe across an imaginary screen in front of you (SIF)", "Bow (B)", "Touch your toes (TYT)".
2. **Lower-Body Motions**: "Bend your knee (BYK)", "Lift your leg to one side (LYL)", "Kick a ball as hard as you can (KBH)".
3. **Full-Body Motions**: "Do a forward lunge (DFL)", "Jump (J)".

We categorized a motion as upper-body if only the limbs of the upper-body are required to perform the motion. Lower-body if only limbs of the lower-body are required to perform the motion. Full-body if both the limbs of the upper and lower body are required to perform the motion.

6.1.3.1 Feature computation

Because we are comparing child and adult motion, we account for height differences: we use participant height as a normalization factor for the global-level features, estimated as the absolute difference between a user's head and foot in the y-dimension of the first frame when performing the Raise hand motion. This movement guarantees that the participant is standing in

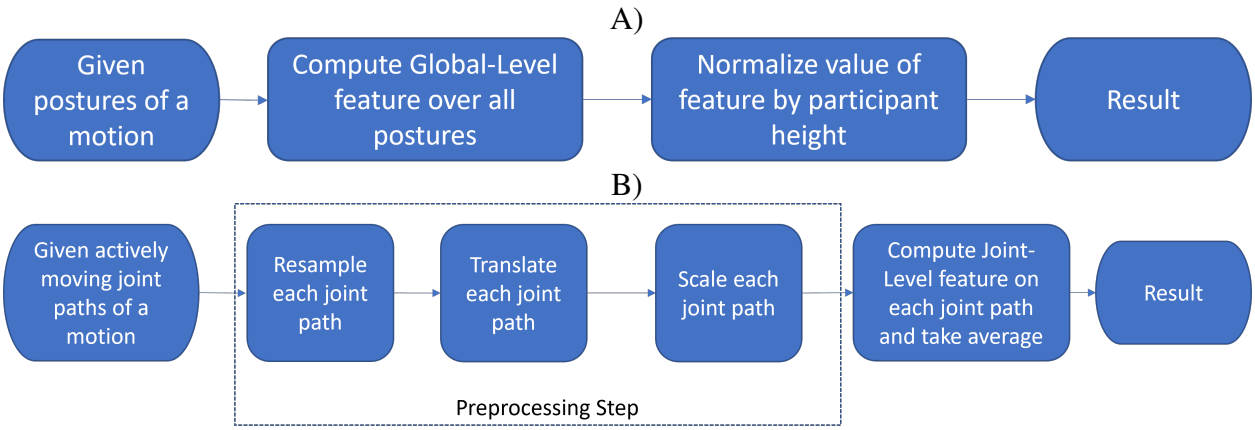


Figure 6-3. Flowcharts showing steps for computing global-level and joint-level features. The motions have already been smoothed using an exponential moving average filter. A) Flowchart detailing the steps for computing global-level features. B) Flowchart detailing the steps for computing joint-level features.

the first frame. For the joint-level features, we account for height differences by scaling the path using the uniform approach proposed in gesture recognition literature [134]. We scaled the path after smoothing the joint articulation path using an exponential moving average filter to remove tracking noise, similar to prior work [5], resampled to $n = 32$ points and translated such that the path's centroid is at the origin. Figures 6-3A and 6-3 are flowcharts showing the steps for computing the global-level and joint-level features, respectively.

6.1.3.2 Statistical analysis

To understand differences between how children and adults perform motions, we analyzed the computed feature values statistically using ANOVA. Because none of our data satisfied the requirements for normality, we used a non-parametric version of the ANOVA test, known as the Aligned Rank Transform (ART) [146]. For each of the features, we ran a two-way repeated-measures ANOVA with a between-subjects factor of age group (child, adult) and a within-subjects factor of motion type (14 motions: touch your toes, point at the camera, raise your hand, raise your arm to one side, bend your knee, put your hands on your hip and lean to one side, do a forward lunge, lift your leg to one side, jump, throw a ball as far as you can, and kick a ball as hard as you can). Since we want to understand how children's motions differ from adults' motions, we are only interested in the main effect of age group and the interaction effect between

age group and motion type (different motion types will be guaranteed to exhibit different feature values). All post-hoc analysis was done using the Tukey method.

6.1.4 Results

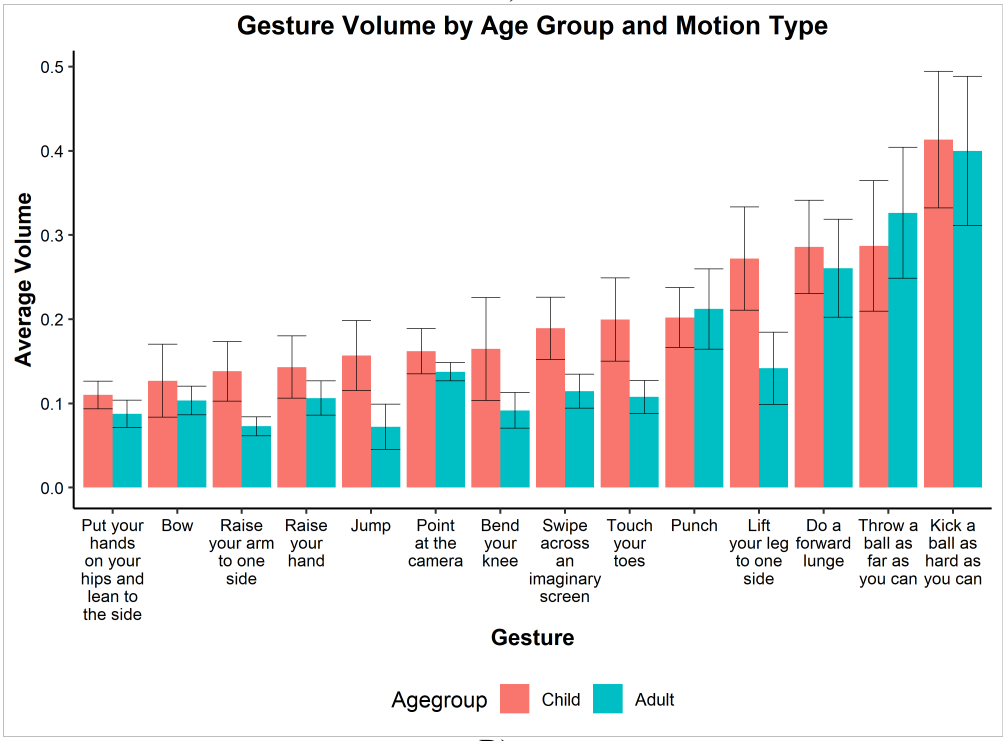
We now present the results from computing the global-level and joint-level features on a subset of children's and adults' motions from the Kinder-Gator dataset [3] and analyzing the results using an ANOVA.

6.1.4.1 Global-level features

As a reminder, we use 13 features defined by Vatavu [132] for characterizing human motion performance comprising spatial features (relates to distance), kinematic features (relates to time), and appearance features (relates to the composition of postures). We present our results below:

Gesture volume (GV) and Gesture area (GA): GV measures the volume of the 3D space where the motion is performed and is computed as the product of the difference between the maximum and minimum positions of the body in the x, y, and z dimensions. GA measures the area of the 2D space in front of the motion sensor and is computed the same way as GV without the z dimension [132]. We found a significant main effect of age group for both features (GV: $F_{1,18} = 6.44, p < 0.05$, GA: $F_{1,18} = 20.18, p < 0.001$), with children requiring a larger 3D and 2D space (GV: mean(M) = $0.20m^3$, standard deviation (SD) = 0.11, median = 0.18, GA: M = $0.55m^2$, SD = 0.17, median = 0.52) as compared to adults (GV: M = $0.16m^3$, SD = 0.12, median = 0.12, GA: M = $0.44m^2$, SD = 0.13, median = 0.41), when normalized for height. We also found a significant interaction effect between age group and motion type for both features (GV: $F_{13,234} = 3.28, p < 0.001$, GA: $F_{13,234} = 4.68, p < 0.001$). Post-hoc tests for GV revealed that children and adults require similar 3D space for all motion types except "Jump", "Lift your leg to one side", and "Touch your Toes", wherein children require a larger 3D space (Figure 6-5A). For GA, children required a larger 2D space for all motion types except "Jump", "Swipe across an imaginary screen", and "Raise your hand", wherein children required a larger 2D space (Figure 6-5B). Therefore, in general, children use more space (proportionally) compared to adults.

A)



B)

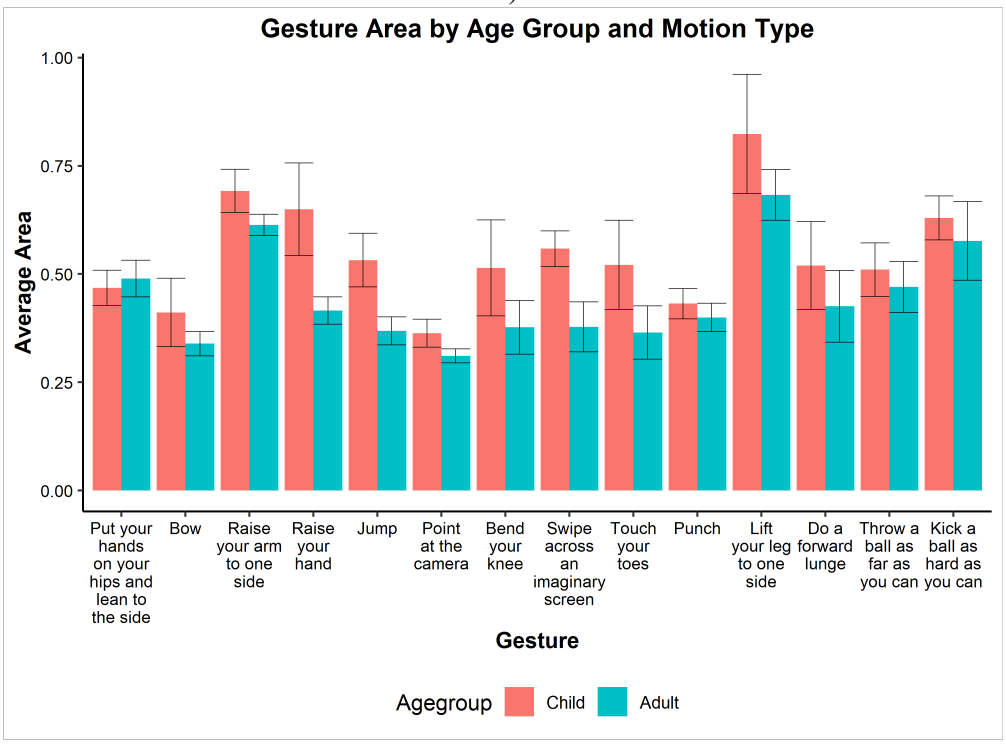
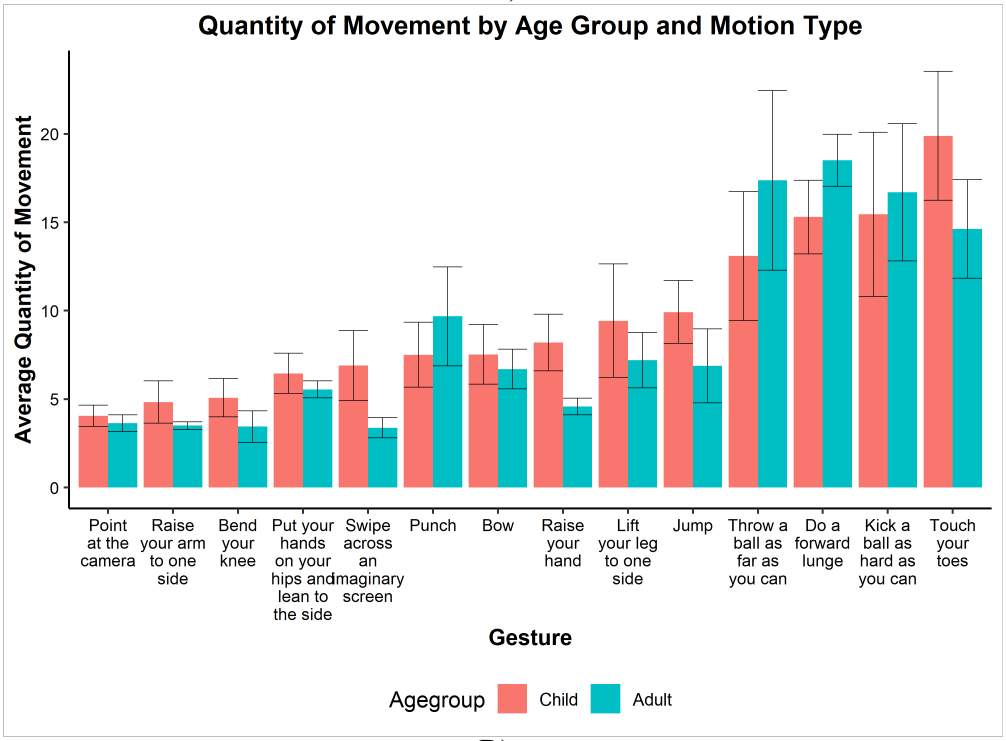


Figure 6-4. Gesture Volume and Gesture Area by motion type and age group. A) Gesture volume by motion type and age group, ordered in ascending order of children's gesture volume. B) Gesture area by motion type and population. Error bars denote 95% confidence interval.

Quantity of movement (QM), Quantity of upper body movement (QUM), Quantity of lower body movement (QLM): QM, QUM, and QLM measure the amount of movement in the whole body, upper body, and lower body, respectively and is computed as cumulative pairwise Euclidean distance between corresponding joints of time–consecutive frames of the motion. We found no significant main effect of age group for all features (QM: $F_{1,18} = 0.64, n.s.$, QUM: $F_{1,18} = 1.28, n.s$, QLM: $F_{1,18} = 0.0004, n.s.$). However, we found a significant interaction effect between age group and motion type across all features (QM: $F_{13,234} = 4.79, p < 0.0001$, QUM: $F_{13,234} = 4.20, p < 0.0001$, QLM: $F_{13,234} = 5.22, p < 0.0001$). Post hoc tests for QM revealed that children and adults require a similar quantity of movement for all motions except "Swipe across an imaginary screen" and "Raise your hand" (Figure 6-5A). For QUM, children required more quantity of upper body movement for the motion "Swipe across an imaginary screen" (Figure 6-5B) while for QLM, children required more quantity of lower body movement for the motion "Raise your hand" (Figure 6-6). These findings suggest that for the motion types we considered, children make the same amount of movements as adults. However, children's movements tend to differ from adults when raising their hand or swiping across an imaginary screen.

Difference of movement (DM) and Ratio of movement (RM): DM measures the absolute difference between the upper body and lower body movement while RM measures the ratio of the upper body to lower body movement [132]. We found no significant effect of age group for DM $(F_{1,18} = 3.52, n.s.)$ but found a significant effect of age group for RM: $(F_{1,18} = 35.82, p < 0.001)$, , with children having a lower ratio of upper body to lower body movement (M = 2.91, SD = 2.04, median = 2.41) compared to adults (M = 3.77, SD = 3.45, median = 2.29); children moved their upper body more in comparison to their lower body as compared to adults. We found a significant interaction effect between age group and motion type (DM: $(F_{13,234} = 2.58, p < 0.01)$, RM: $(F_{13,234} = 7.67, p < 0.001)$). For DM, post-hoc tests only revealed a marginal difference between children and adults for the "Swipe across an imaginary screen in front of you" motion (Figure 6-7A). Although we found a significant interaction effect for RM, post-hoc tests revealed no significant difference between interaction pairs after Tukey correction (Figure 6-7B). The Tukey

A)



B)

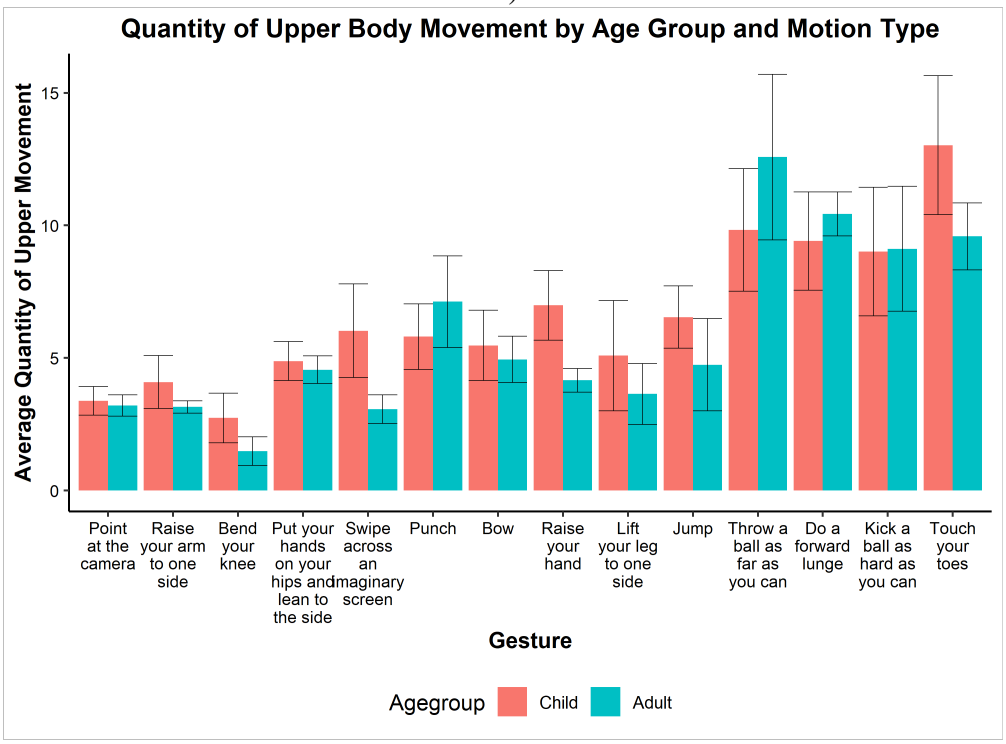


Figure 6-5. Quantity of Movement and Quantity of Upper Body Movement by motion type and age group. A) Quantity of Movement by motion type and population, sorted in ascending order of children's quantity of movement. B) Quantity of Upper Body Movement by motion type and age group. Error bars denote 95% confidence interval.

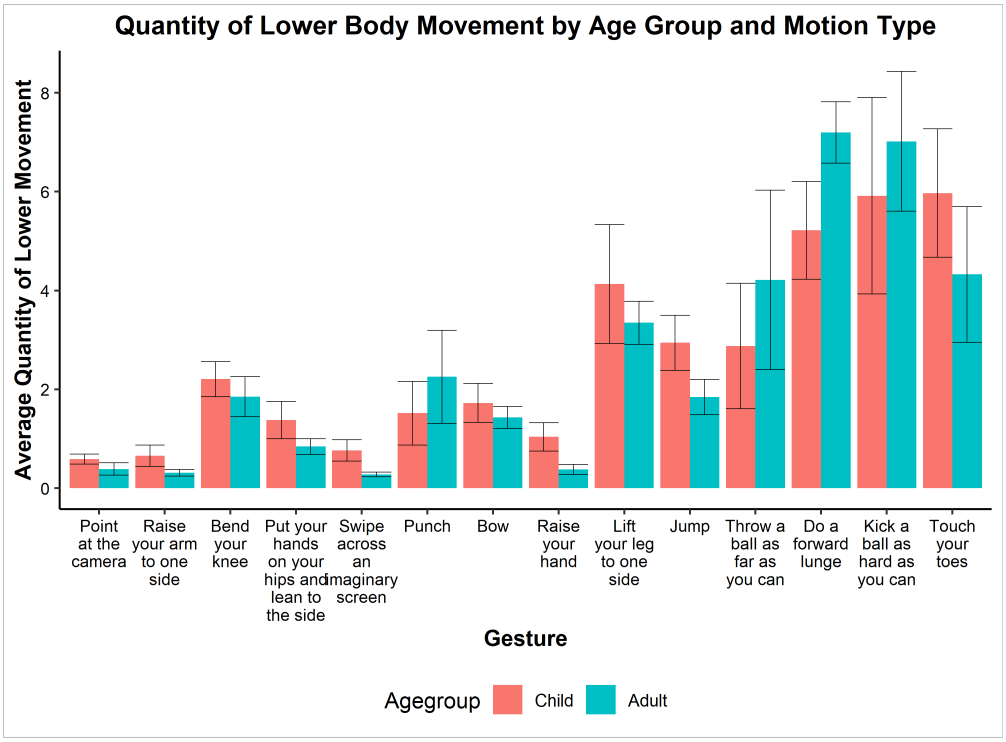
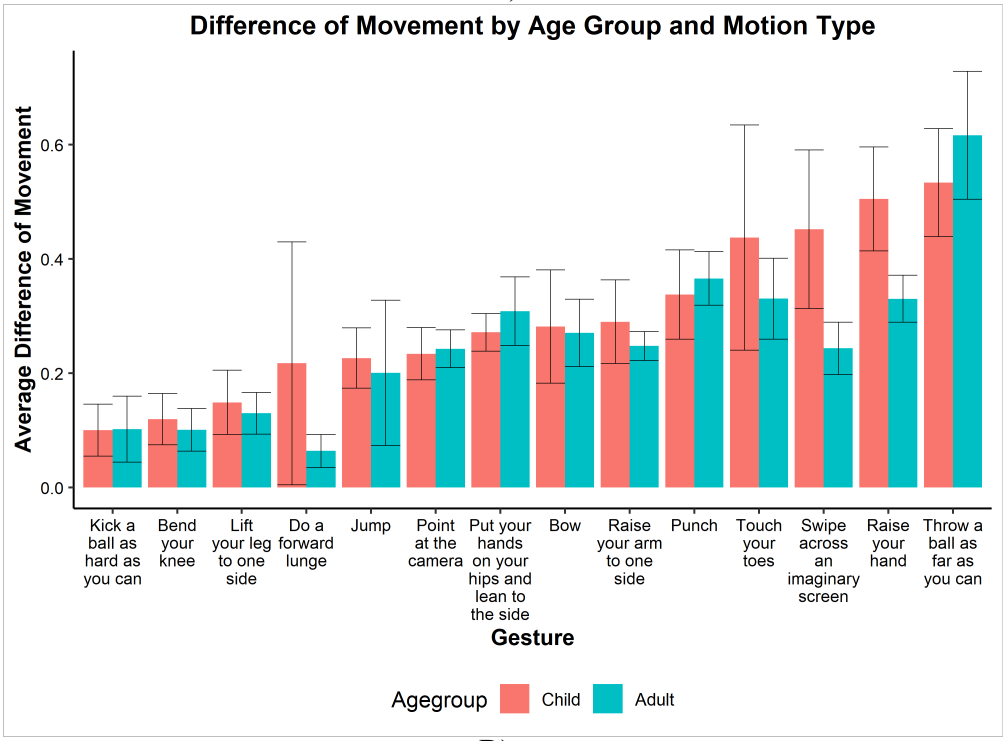


Figure 6-6. Quantity of Lower Body Movement by motion type and age group, order of motion types is same as the figures in Figure 6-5. Error bars denote 95% confidence interval.

post-hoc correction could have significantly penalized the p-values to reduce the risk of type I error (i.e., rejecting the null hypothesis when the hypothesis is true). These findings indicate that for the motion types we considered, children move their upper body more in comparison to their lower body.

Performance time (PM) and Average gesture speed (AGS): PT measures the time a user takes to perform the motion while AGS is the ratio of quantity of movement to performance time [132]. We found a significant main effect of age group for both features (PT: $(F_{1,18} = 8.08, p < 0.01)$, AGS: $(F_{1,18} = 29.37, p < 0.001)$) with children moving faster (M = 3.20s, SD = 1.17, median = 2.90) and at a higher speed (M = 0.16m/s, SD = 0.09, median = 0.14) as compared to adults (PT: m = 3.76s, SD = 1.19, median = 0.14, AGS: M = 0.12m/s, SD = 0.07, median = 0.09). We also found a significant interaction effect between age group and motion type for both features (PT: $(F_{13,234} = 2.26, p < 0.01)$, AGS: $(F_{13,234} = 2.83, p < 0.01)$). Post-hoc tests revealed that children moved at similar speeds as adults for all motion types except "Bend your

A)



B)

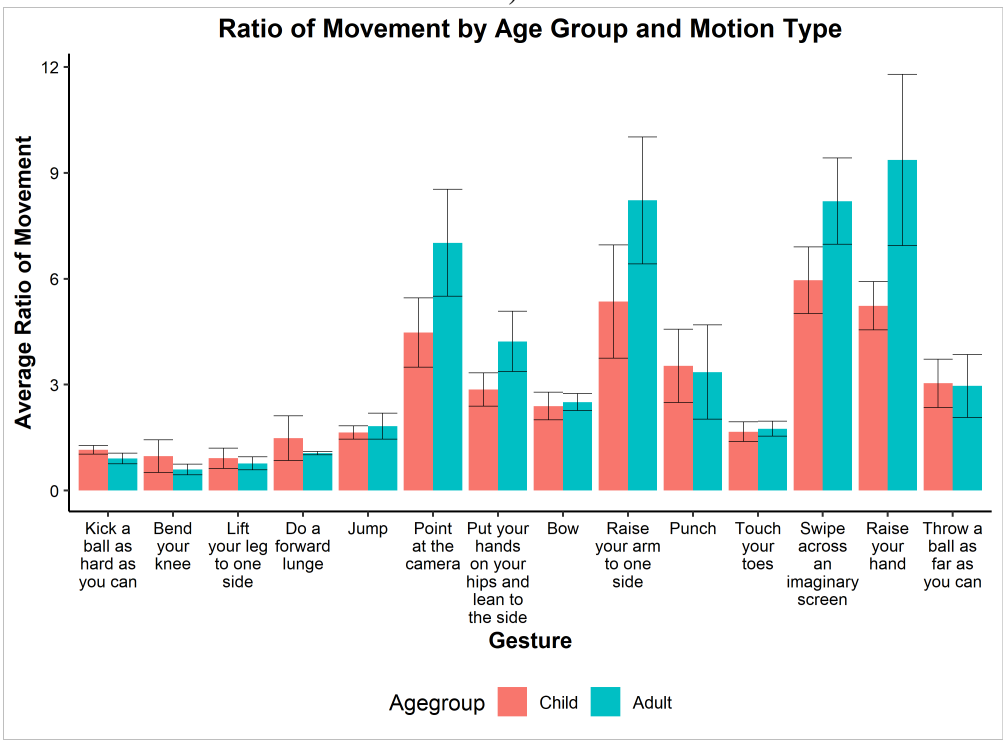


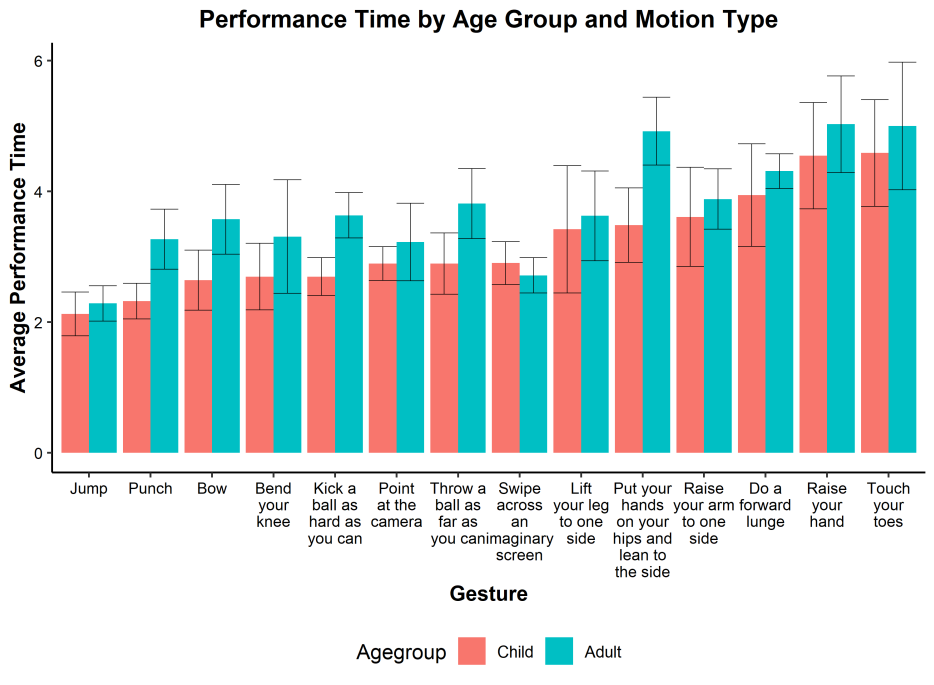
Figure 6-7. Difference of Movement and Ratio of Movement by motion type and population. A) Difference of Movement by motion type and population, sorted in ascending order of children's difference of movement. B) Ratio of Movement by motion type and population. Error bars denote 95% confidence interval.

knee", "Jump", "Raise your hand", and "Swipe across an imaginary screen" (Figure 6-8A). Like ratio of movement, post-hoc tests for performance time showed no significant difference between interaction pairs after Tukey correction (Figure 6-8B). These findings indicate that children generally move faster than adults, corroborating findings from prior work [4].

Body posture variation (BPV) and Body posture diffusion (BPD): BPV measures the deviation of postures from the centroid posture while BPD measures the maximum difference between postures [132]. We found no significant effect of age group for both features (BPV: $(F_{1,18} = 0.67, n.s.)$, BPD: $(F_{1,18} = 0.95, n.s.)$). However, we found a significant interaction effect (BPV: $(F_{13,234} = 80.88, p < 0.001)$, BPD: $(F_{13,234} = 4.18, p < 0.001)$). Children and adults had similar posture variations for all motion types except "Raise your hand" and "Jump", wherein children have a larger body posture variation than adults (Figure 6-9A). Children and adults have similar posture diffusion for all motion types except "Raise your hand", wherein we found a marginal difference (Figure 6-9B). These findings suggest that children and adults use similar posture compositions to perform motions.

Body posture density (BPDe) and Body posture rate (BPR). BPDe is the ratio of Body posture variation (BPV) to gesture volume (GV) while BPR is the ratio of BPV to performance time [132]. We found a significant main effect of age group (BPDe: $(F_{1,18} = 10.74, p < 0.01)$, BPR: $(F_{1,18} = 18.99, p < 0.01)$), with children having lower posture densities (M = 6.70m^{-2}, SD = 3.19, median = 5.89) but higher posture rates (M = 0.42m/s, SD = 0.26, median = 0.37) as compared to adults (BPDe: M = 8.05m^{-2}, SD = 4.28, median = 6.87, BPR: M = 0.32m/s, SD = 0.21, median = 0.24). We also found a significant interaction effect between age group and motion type for both features (BPDe: $(F_{13,234} = 3.16, p < 0.001)$, BPR: $(F_{13,234} = 1.91, p < 0.05)$). Children and adults had similar posture densities for all motion types except "Lift your leg to one side" (Figure 6-10A) and had similar posture rates for all motion types except "Bend your knee", "Jump", and "Put your hands on your hips" (Figure 6-10B). Vatavu [132] noted that both of these features help to differentiate between motions that have a similar Body posture variation. These findings suggest that children's and adults' motions will

A)



B)

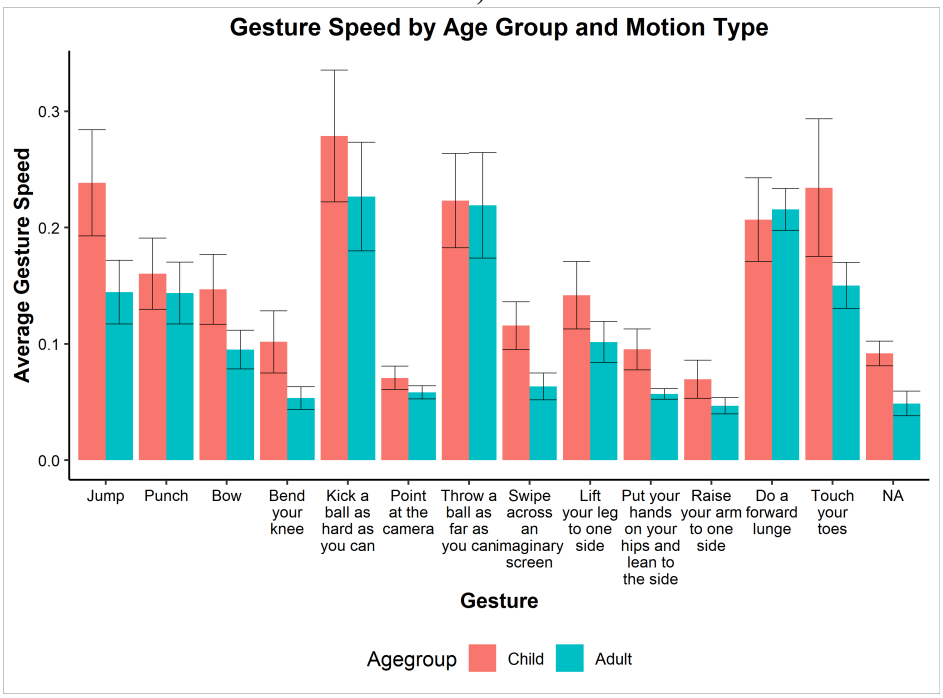
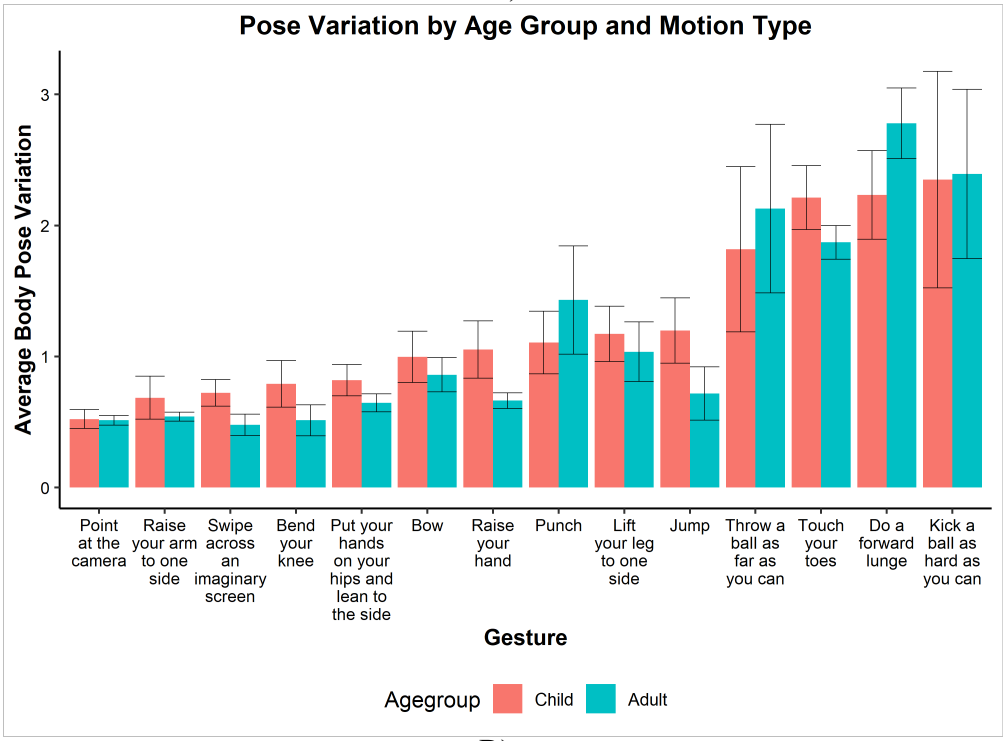


Figure 6-8. Performance Time and Average Gesture Speed by motion type and population. A) Performance Time by motion type and population, sorted in ascending order of children's performance time. B) Average Gesture Speed by motion type and population. Error bars denote 95% confidence interval.

A)



B)

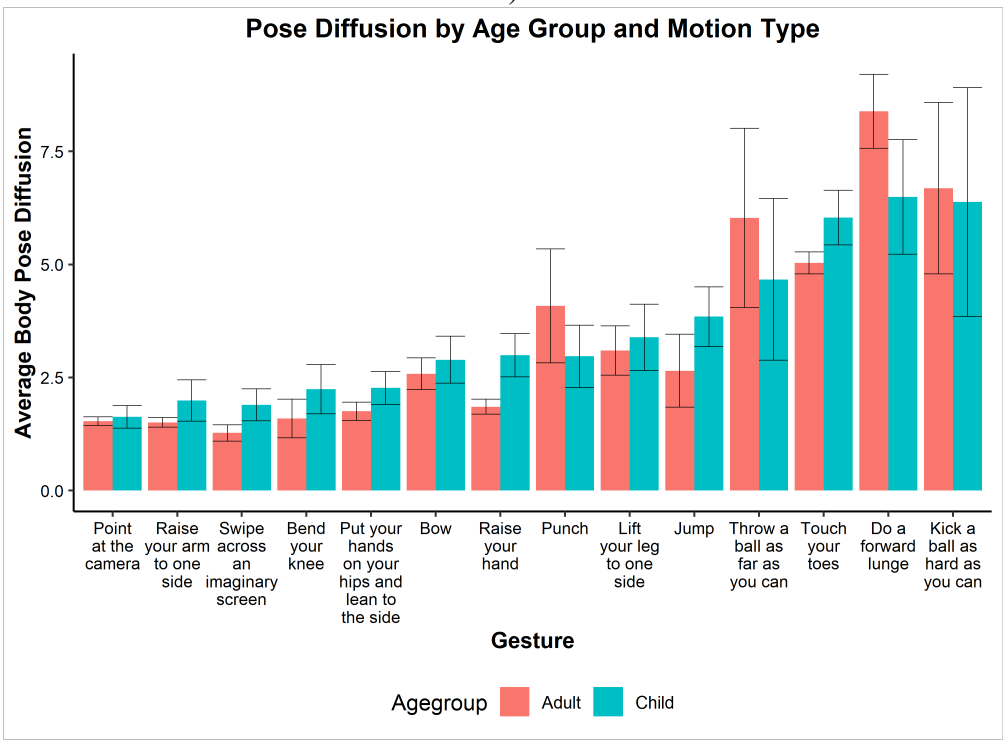


Figure 6-9. Body Posture Variation and Body Posture Diffusion by motion type and population. A) Body Posture Variation by motion type and population, sorted in ascending order of children's body posture variation. B) Body Posture Diffusion by motion type and population. Error bars denote 95% confidence interval.

differ in appearance (i.e., the distribution of their body postures) when the space and time they require to perform the motion are considered.

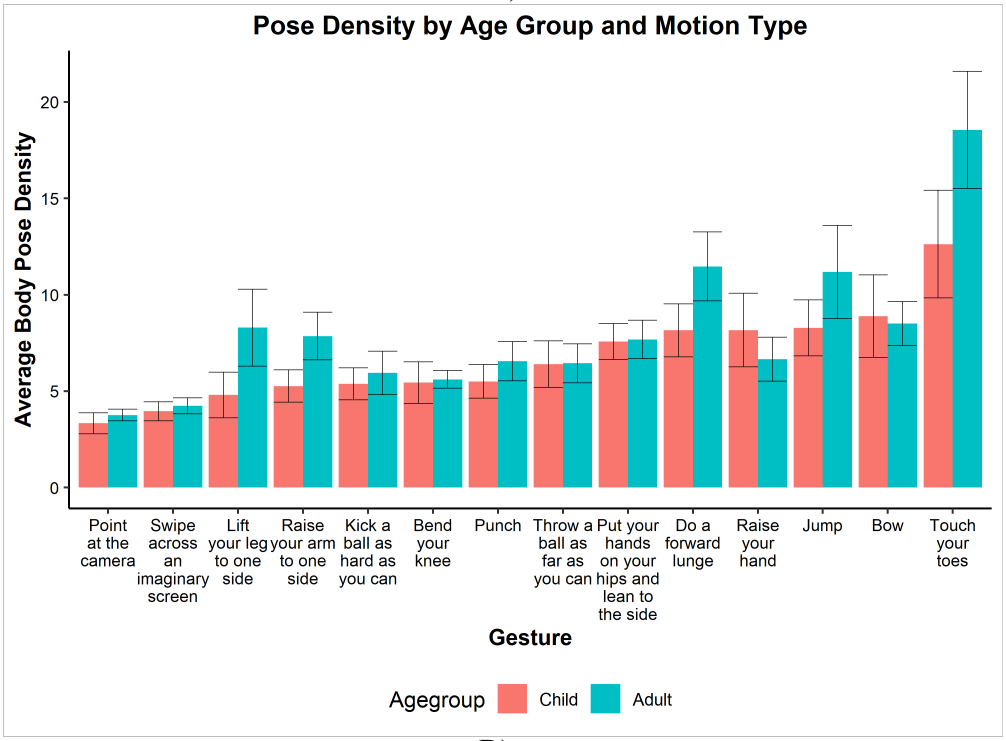
6.1.4.2 Joint-level Features

Next, we report findings from computing our 11 geometric features on the same subset of 14 children's and adults' motions. Similar to the results presented above, we are interested in the main effect of age group (child, adult) and interaction effect between age group and the motion types.

Shape error (ShE) and Shape variability (ShV). ShE measures the average absolute deviation of the shape of a joint path in a motion instance from the shape of the same joint path in the joint task while ShV measures how uniform the shape errors are along a joint path. We found a significant effect of age group for both features (ShE: $(F_{1,18} = 137.20, p < 0.0001)$, ShV: $(F_{1,18} = 25.62, p < 0.0001)$) with children having higher shape errors (M = 0.31, SD = 0.08, median = 0.31) and shape variabilities (M = 0.17, SD = 0.06, median = 0.17) compared to adults (ShE: M = 0.20, SD = 0.08, median = 0.21, ShV: M = 0.11, SD = 0.06, median = 0.10). We found a significant interaction effect between age group and motion type for both features (ShE: $(F_{13,234} = 4.70, p < 0.0001)$, ShV: $(F_{13,234} = 4.41, p < 0.0001)$). Post hoc tests revealed that children and adults have similar shape errors for all motion types, except "Bend your knee", "Bow", "Do a forward lunge", "Jump", "Lift your leg to one side", "Point at the camera", "Put your hands on your hips and lean to the side", "Raise your arm to one side", "Swipe across an imaginary screen", and "Throw a ball as far as you can", for which chidren have higher shape errors (Figure 6-11A). Children had higher shape variability compared to adults for all motion types, except for "Do a forward lunge", "Kick a ball as hard as you can", "Point at the camera", "Swipe across an imaginary screen", and "Throw a ball as far as you can" (Figure 6-11B). Therefore, children show more variations in how they move their body parts as compared to adults for most motion types, thus making them more inconsistent.

Bend error (BE) and Bend variability (BV). BE measures users' tendency to bend or curve the joint path while BV measures how uniform the bend errors are along the joint path. We

A)



B)

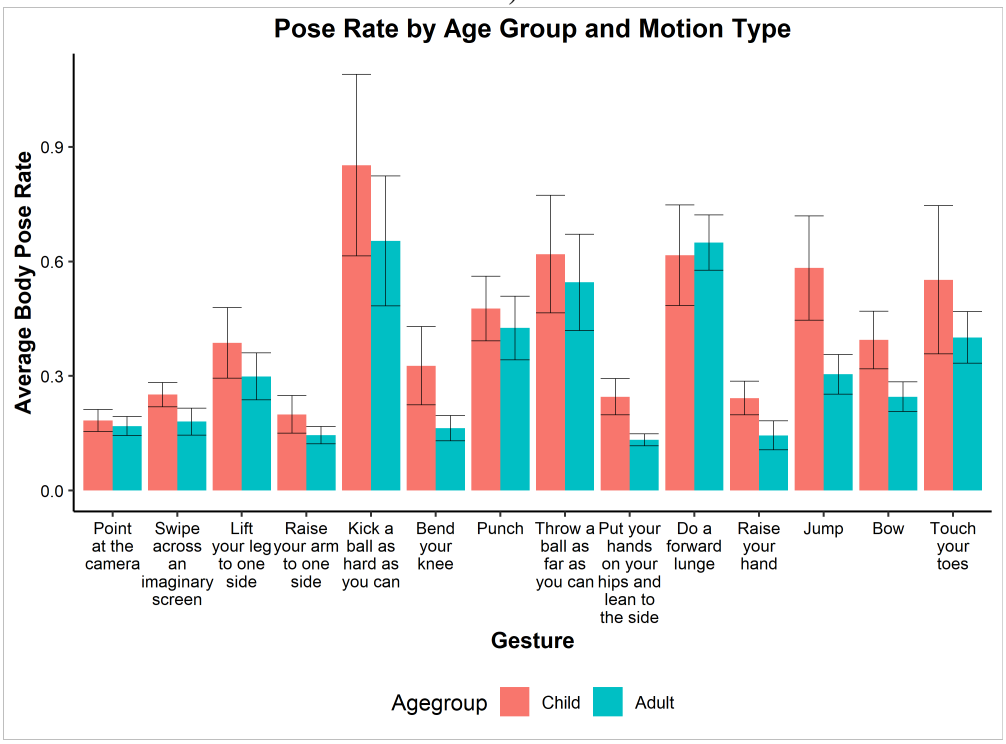
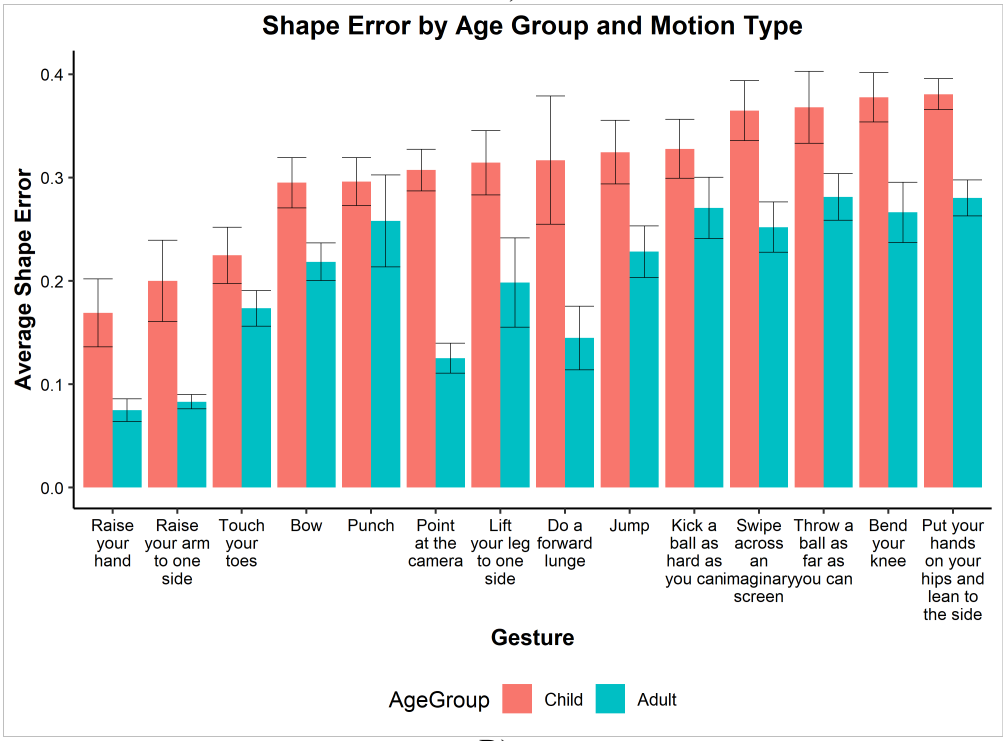


Figure 6-10. Body Posture Density and Body Posture Rate by motion type and population. A) Body Posture Density by motion type and population, sorted in ascending order of children's body posture density. B) Body Posture Rate by motion type and population. Error bars denote 95% confidence interval.

A)



B)

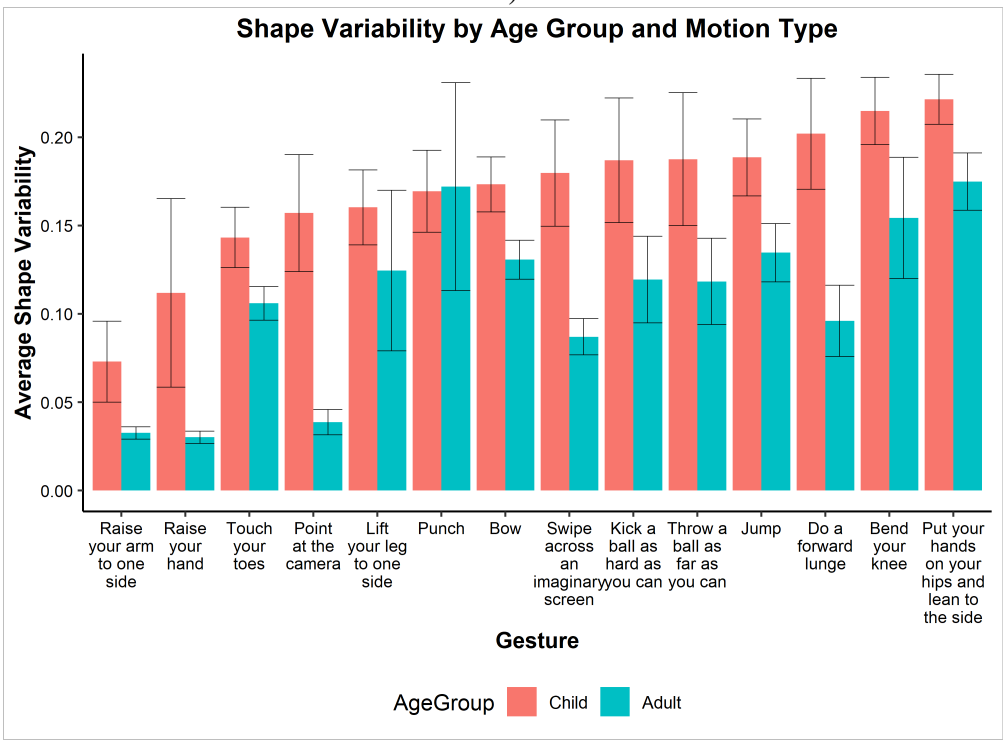
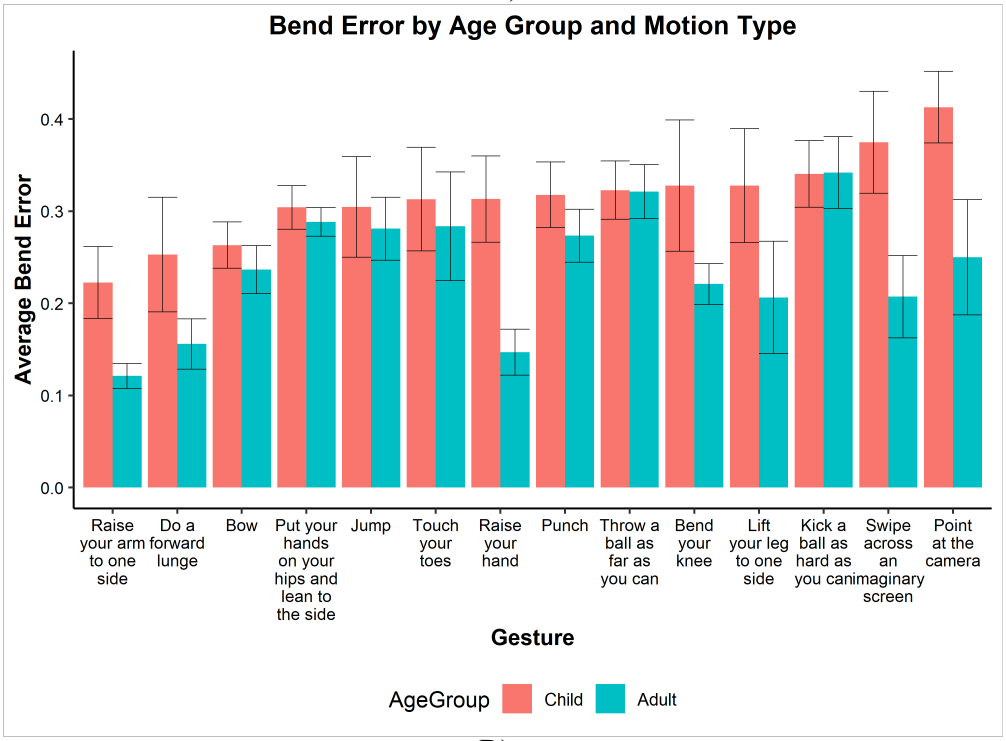


Figure 6-11. Shape Error and Shape Variability by motion type and population. A) Shape Error by motion type and population, sorted in ascending order of children's shape errors. B) Shape Variability by motion type and population. Error bars denote 95% confidence interval.

found a significant main effect of age group for both features (BE: $F_{1,18} = 78.61, p < 0.0001$, BV: $F_{1,18} = 36.50, p < 0.0001$) with children having higher bend errors (M = 0.31, SD = 0.09, median = 0.30) and bend variabilities (M = 0.35, SD = 0.10, median = 0.36) as compared to adults (BE: M = 0.24, SD = 0.09, median = 0.23, BV: M = 0.41, SD = 0.09, median = 0.39). We also found a significant interaction effect between age group and motion type (BE: $F_{13,234} = 7.35, p < 0.0001$, BV: $F_{13,234} = 6.81, p < 0.0001$). Post hoc tests revealed that children and adults have similar bend errors for all motion types, except "Lift your leg to one side", "Point at the camera", "Raise your hand", and "Swipe across an imaginary screen" (Figure 6-12A). Children also had higher bend variability compared to adults for all motion types, except for "Raise your arm to one side", "Raise your hand", and "Swipe across an imaginary screen" (Figure 6-12B). Hence, children vary in how they bend their body parts to perform motions across instances of the same motion.

Length error (LE) and Size error (SE). LE measures a user's tendency to stretch their joint path, as measured by the path length, while SE measures a user's tendency to stretch the joint path with respect to the gesture volume (GV). We found a significant main effect of age group for both features (LE: $(F_{1,18} = 35.28, p < 0.0001)$, SE: $F_{1,18} = 37.44, p < 0.0001$) with children having higher length errors (M = 0.44, SD = 0.27, median = 0.39) and higher size errors (M = 0.13, SD = 0.06, median = 0.13) as compared to adults (LE: $F_{13,234} = 6.42, p < 0.0001$, EF: $(F_{13,234} = 3.21, p < 0.0001)$). We also found a significant interaction effect between age group and motion type for both features (LE: $F_{13,234} = 6.42, p < 0.0001$, SE: $(F_{13,234} = 3.21, p < 0.0001)$). Post-hoc tests revealed that children and adults have similar length errors for all motions except "Lift your leg to one side", "Point at the camera", "Raise your hand", and "Do a forward lunge" (Figure 6-13A). Children had higher size errors than adults for the "Do a forward lunge" motion (Figure 6-13B). The length error results corroborate the shape error results in that children are inconsistent in how they move their body parts. The size error results indicate that not only do children require more space to perform motions (see gesture volume and gesture area results), but they also vary in the amount of space they require to perform motions across instances of the same motion.

A)



B)

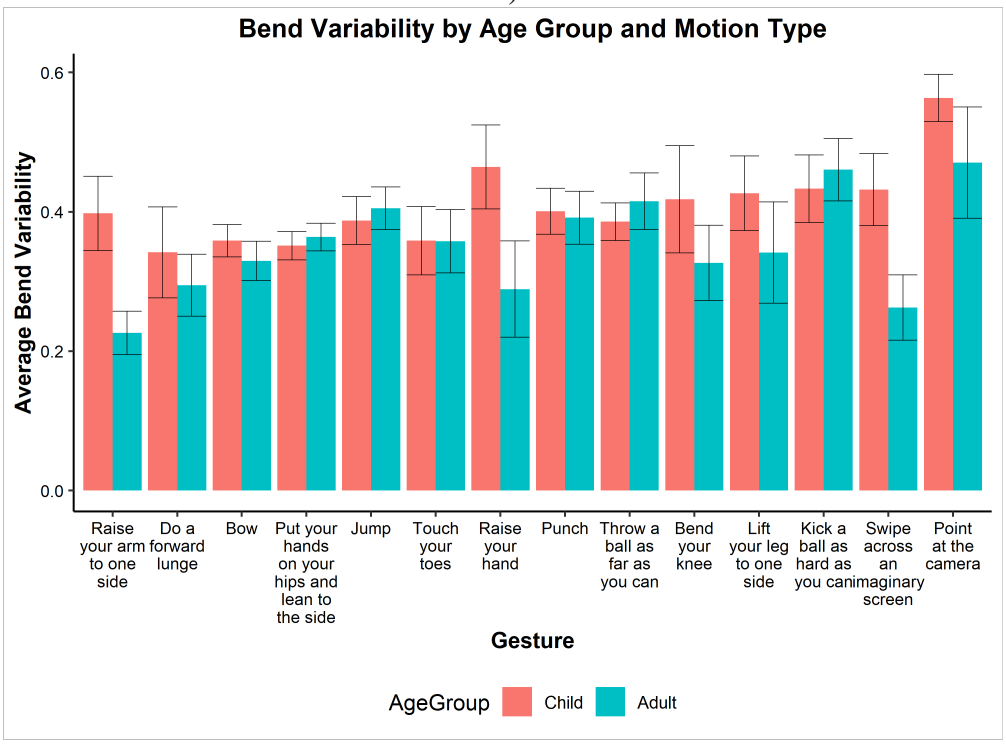
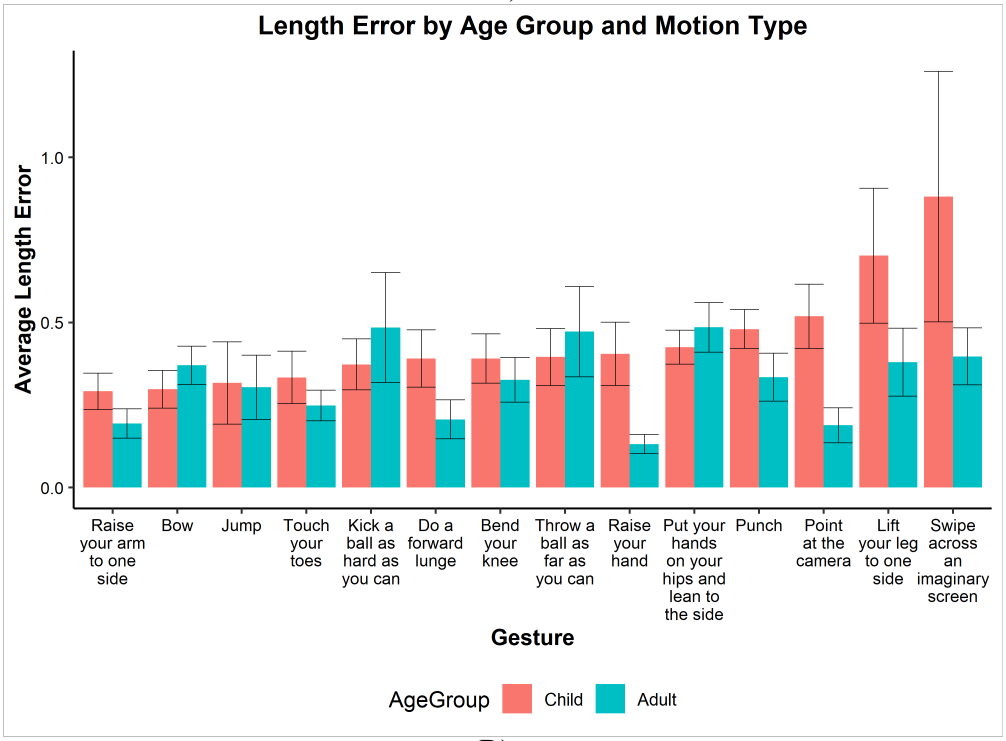


Figure 6-12. Bend Error and Bend Variability by motion type and population. A) Bend Error by motion type and population, sorted in ascending order of children's bend errors. B) Bend Variability by motion type and population. Error bars denote 95% confidence interval.

A)



B)

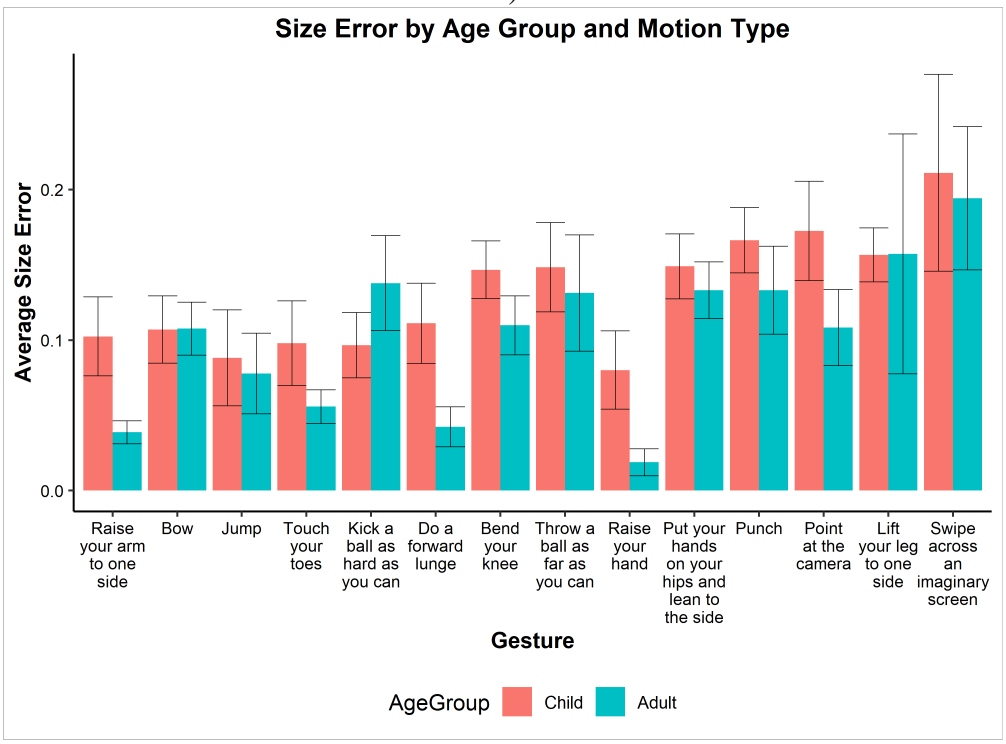


Figure 6-13. Length Error and Size Error by motion type and age group. A) Length Error by motion type and age group, sorted in ascending order of children's length errors. B) Size Error by motion type and age group, , sorted in ascending order of children's length errors. Error bars denote 95% confidence interval.

Efficiency (E). Efficiency measures how efficient a path is in comparison to the joint task axis, which is the ideal path. We found a marginal effect of age group for efficiency ($F_{1,18} = 3.93, p = 0.06$) and no significant interaction effect between age group and motion type ($F_{13,234} = 0.53, n.s.$) (Figure 6-14). These findings suggest that for the motion types considered, children move just as efficiently as adults. However, we note that the marginal effect shows the potential of children being less efficient than adults. We posit that age will play a major role in the efficiency of children's motion performance with younger children ($\leq$ 6 years) being less efficient than older children ($>$ 6 years) because their motor abilities are less developed [23].

Time error (TE) and Speed error (VE). TE measures the difference between the total time it takes to articulate a joint path to the time it takes to articulate the joint task axis while VE measures the average difference in the speed of the joint path in a motion instance and the speed of a joint path in the corresponding joint task axis. We found no significant main effect of age group for TE ($F_{1,18} = 0.74, n.s.$), but found a significant interaction effect between age group and motion type ($F_{13,234} = 9.33, p < 0.0001$). Post hoc tests revealed that children and adults have similar time errors for all motion types, except "Do a forward lunge", for which children had a lower time error than adults and "Point at the camera", for which children had a higher time error compared to adults (Figure 6-15A). In contrast to time error, we found a significant main effect of age group for speed error ($F_{1,18} = 23.05, p < 0.001$), with children having a higher speed error (M = 0.30, SD = 0.12, median = 0.27) as compared to adults (M = 0.27, SD = 0.10, median = 0.21). We also found a significant interaction effect ($F_{13,234} = 3.51, p < 0.0001$). Children and adults have similar speed errors for all motion types, except "Do a forward lunge" and "Put your hands on your hips and lean to the side", for which children had higher speed errors as compared to adults (Figure 6-15B). Taken together with performance time (a global-level feature), our time error results suggest that children are consistently faster than adults when performing motions. Our speed error results indicate that even though children move consistently faster than adults, they vary in the rate at which they move their body parts throughout the motion, further corroborating our length error findings.

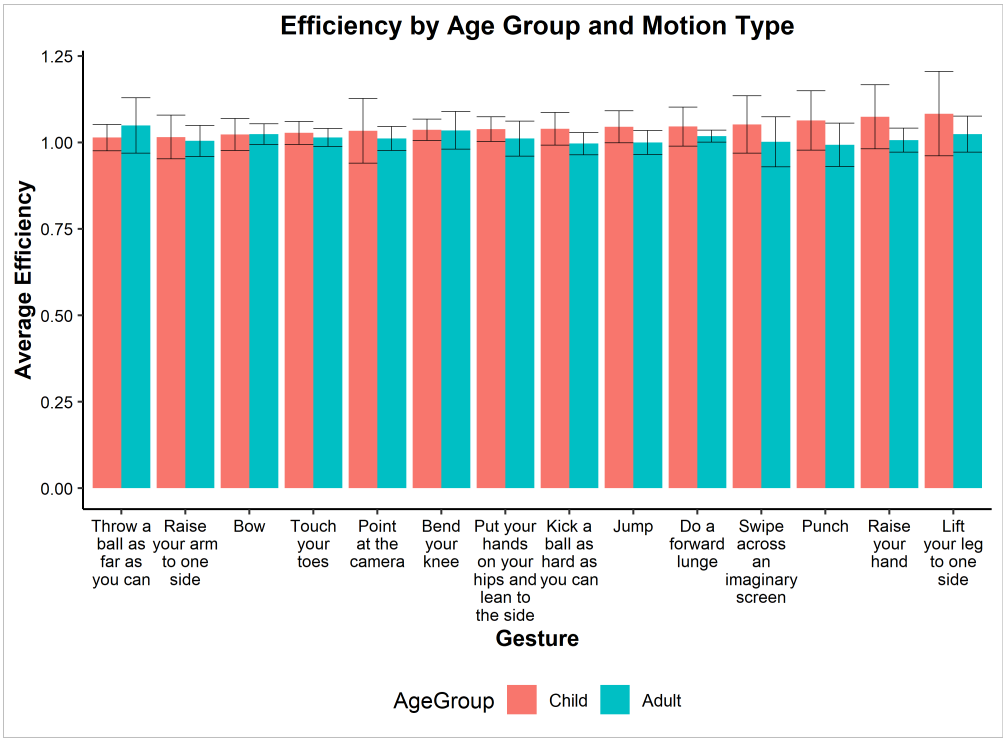
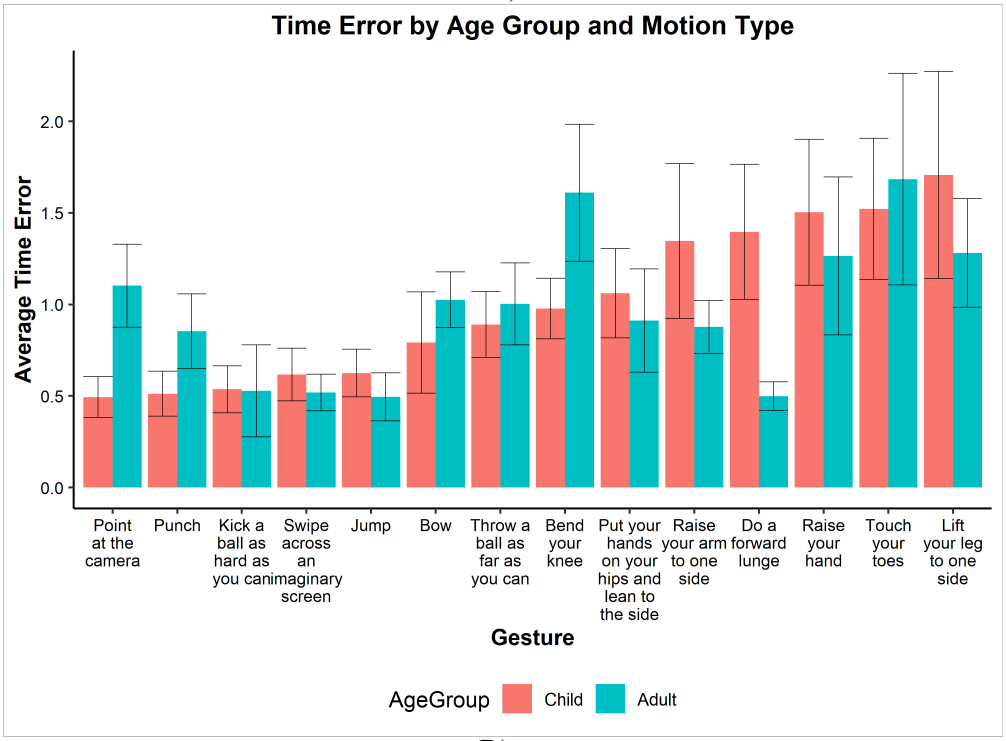


Figure 6-14. Efficiency by motion type and age group sorted in ascending order of children's efficiency errors

Acceleration error (AE) and Jerk error (JE). AE and JE measure the average difference in the acceleration and jerk of a joint path and the acceleration and jerk of a joint path in the corresponding joint task axis respectively. We found a significant main effect of age group for both features (AE: $(F_{1,18} = 80.67, p < 0.0001)$, JE: $F_{1,18} = 70.07, p < 0.0001$), with children having higher acceleration errors (M = 0.28, SD = 0.13, MEDIAN = 0.25) and jerk errors (M = 2.85, SD = 1.72, MEDIAN = 2.41) as compared to adults (AE: M = 0.19, SD = 0.11, MEDIAN = 0.17, JE: M = 1.69, SD = 1.49, MEDIAN = 1.28). We also found a significant interaction effect between age group and motion type for both features (AE: $F_{13,234} = 5.01, p < 0.0001$, JE: $F_{13,234} = 5.92, p < 0.0001$). Post hoc tests revealed that children and adults have similar acceleration errors for all motion types, except "Punch", "Raise your arm to one side", and "Swipe across an imaginary screen", for which children had higher acceleration errors as compared to adults (Figure 6-16A). Children had higher jerk errors than adults for motions "Bow", "Lift your leg to one side", "Punch", "Put your hands on your hips and lean to one

A)



B)

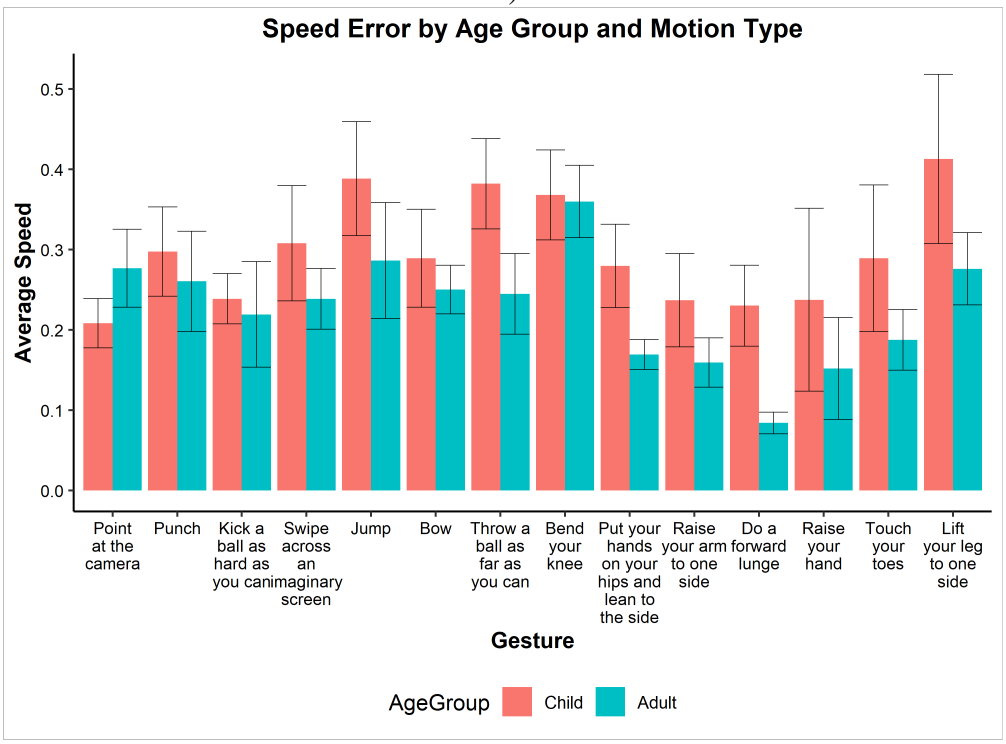


Figure 6-15. Time Error and Speed Error by motion type and population. A) Time Error by motion type and population, sorted in ascending order of children's time errors. B) Speed Error by motion type and population. Error bars denote 95% confidence interval.

side”,”Raise your arm to one side”,”Swipe across an imaginary screen”, and ”Throw a ball as far as you can” (Figure 6-16B). Prior work has noted that high levels of jerk indicate that the motions are performed quickly, with more urgency and with less smoothness [137]. Since the global-level features showed us that children move faster and with higher speeds, taken with these results, we can also deduce that children move with significantly higher acceleration and jerk than adults. Hence, children's motions will be less smooth in general as compared to adults' motions.

For the global-level features, we found a significant difference in 6 of the 13 features, which indicates that children's motions in the Kinder-Gator dataset differed from adults' motions with respect to the specific property being measured by each of these six features (starred in Table 6-1). For the joint-level features, we found significant differences in all but 2 of the 11 features (starred in Table 6-1). Since the joint-level features compare articulation paths, this finding indicates that children move more inconsistently than adults for each of the specific property being measured.

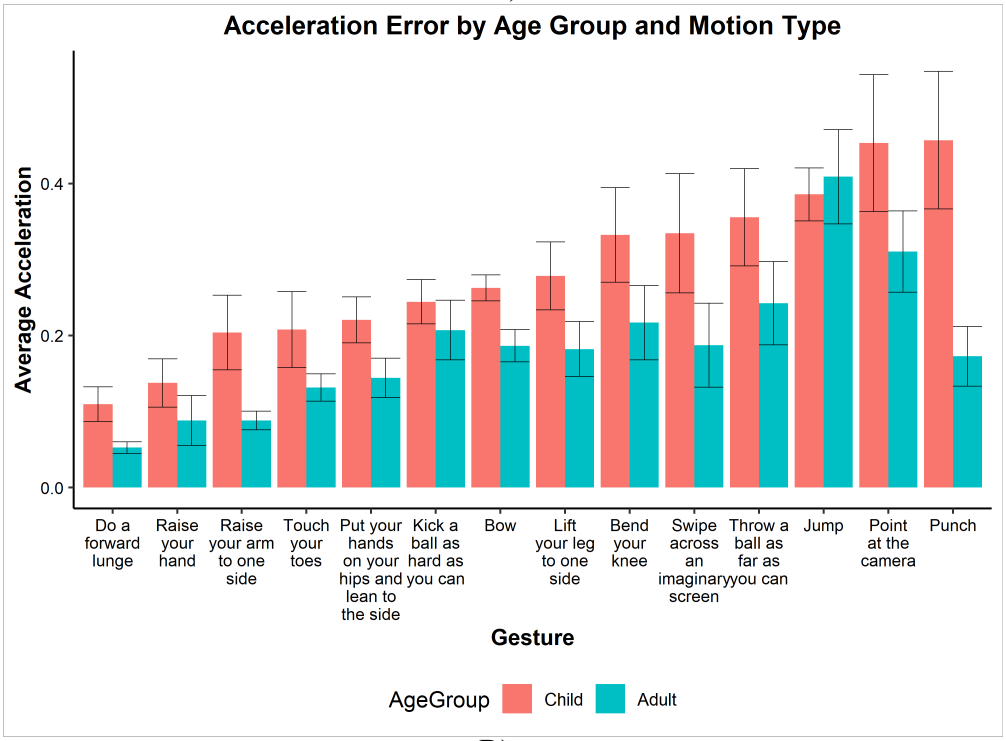
6.1.5 Feature Validation

Although ANOVA tests are useful for identifying features that are relevant for distinguishing between children and adults, this test might not capture the optimal subset of features that distinguish child motion from adult motion because it does not consider feature interactions [151]. For example, ANOVA test can tell us whether age group has an impact on a feature but does not consider the impact of age group on multiple features for its analysis. Hence, in addition to the statistical test above, we apply a feature selection method to identify the optimal subset of features for differentiating child motion from adults motions. Specifically, we use the Wrapper method [84].

6.1.5.1 Wrapper method

The Wrapper method [84] uses a machine-learning classifier as a “black box” to identify a subset of features that results in the best performance for that classifier. Although Wrapper methods are more computationally expensive than ANOVAs, which are also used for feature selection, their results are more accurate [84]. Therefore, we use our ANOVA results as a pre-selection step prior to identifying the optimal subset of features. For the Wrapper method, we

A)



B)

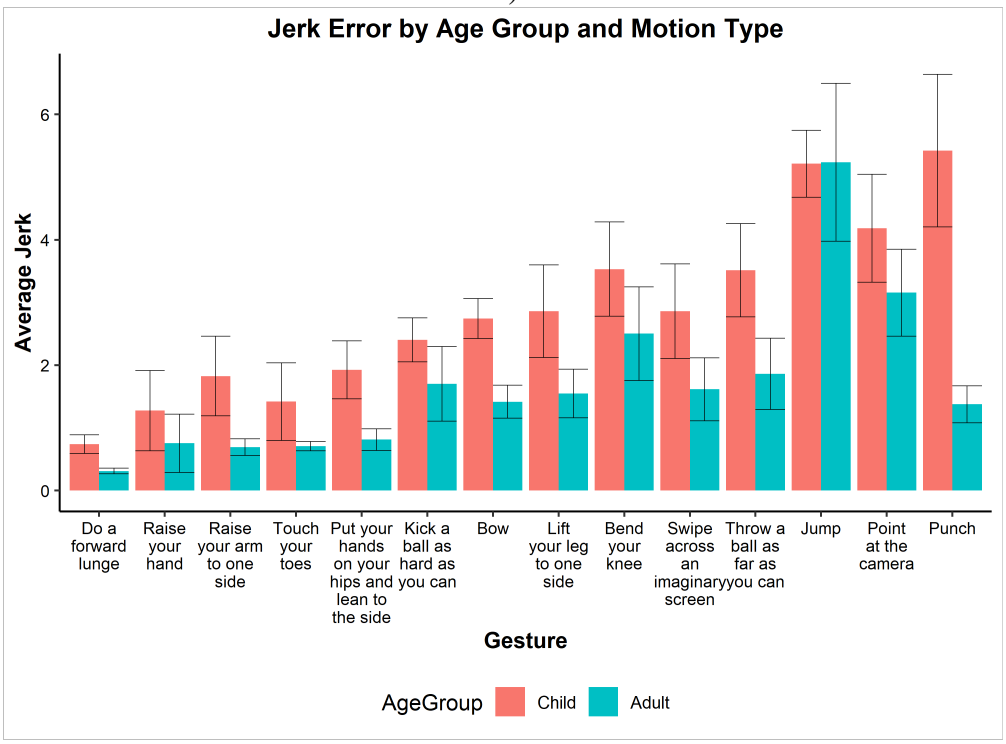


Figure 6-16. Acceleration Error and Jerk Error by motion type and population. A) Acceleration Error by motion type and population, sorted in ascending order of children's acceleration errors. B) Jerk Error by motion type and population. Error bars denote 95% confidence interval.

Table 6-1. List of Articulation Features. * means the feature was significant with respect to age group at $p < 0.05$.

Global-Level Features (13)	Joint-Level Features (11)
Gesture Volume*	Shape Error*
Gesture Area*	Shape Variability*
Quantity of Body Movement	Bend Error*
Quantity of Upper Body Movement	Bend Variability*
Quantity of Lower Body Movement	Length Error*
Difference of Movement	Size Error*
Ratio of Movement*	Efficiency
Performance Time*	Time Error
Average Gesture Speed*	Speed Error*
Body Posture Variation	Acceleration Error*
Body Posture Diffusion	Jerk Error*
Body Pose Density*	
Body Pose Rate*	

use the Holdout-SVM linear kernel Wrapper method algorithm (HO-SVM) proposed by Maldonado and Weber [84]. This algorithm combines an SVM classifier with the backward selection pruning method to select an optimal subset of features. In backward selection, the algorithm begins with the full set of features and then recursively removes the feature without which the classifier achieves the best performance [84]. Consequently, Wrapper methods are also known Recursive Feature Elimination (RFE) methods [47]. To implement the HO-SVM algorithm, we directly apply the steps described in Maldonado's and Weber's paper [84] with some modifications:

Model selection. Determine the parameters for the SVM. Because we are using a linear kernel, the only parameter we need to determine is C, which determines the margin of the hyperplane. We used a Receiver Operating Characteristic (ROC) Curve with varying values of C = $\{0.1, 0.5, 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$. C was set to 20.

Initialization. Initialize a feature set α comprising the full set of features (Backward Selection). Because we are using the ANOVA results as a pre-selection phase, the full set of features will only be the features in which we found a significant effect of age group, namely: (all global-level features except quantity of movement, quantity of upper body movement, quantity of lower body movement, body pose variation and body pose diffusion and all joint-level features

except time error and efficiency). Hence, each motion is initially represented in the form (x,y) where x represents a set of articulation features and y represents the true label in which the motion belongs (i.e., child, adult).

Repeat the following five steps until the maximum recognition accuracy achieved after feature removal is less than or equal to the recognition accuracy without removing any of the features in the current feature set (i.e., $max(\Gamma_{\alpha_{(-p)}}) \leq \Gamma_{\alpha}$).:

1. Train and test the SVM using 10-fold cross-validation. In the original algorithm, the authors randomly split the data into training and testing and suggested that we run the entire algorithm multiple times to get the optimal subset. However, we noticed that this idea introduced too much randomness in our results, with each run returning a different set of feature. Hence, we used cross-validation, which has been used extensively in machine learning to guarantee the reliability of the classifier.
2. Compute recognition accuracy Γ_{α}. The original algorithm uses recognition error instead, which is equivalent (accuracy = 1-error).
3. For each feature p in α, remove p from the current set of features.
4. Train and test the SVM classifier on the remaining subset of features using 10-fold cross-validation and compute recognition accuracy $\Gamma_{\alpha_{(-p)}}$.
5. Remove feature p with the maximum accuracy (i.e., $max(\Gamma_{\alpha_{(-p)}})$ from among all features in the current feature set α.

Next, we discuss our findings after applying the Wrapper method to the subset of features that showed a significant effect based on the ANOVA results.

6.1.5.2 Results

Our algorithm terminated after two iterations (i.e., two features were removed from the set of features). In the first iteration, "Average Gesture Speed" was removed while in the second iteration, "Bend Variability" was removed, with a final recognition accuracy of 80.7%. The resulting optimal set of 14 features for distinguishing child motion from adult motion is shown in Table 6-2.

6.1.6 Summary

So far, we have presented a set of global-level features and joint-level features to quantitatively describe motions. Using these features, we analyzed a subset of children's and

Table 6-2. Optimal subset of features after applying ANOVA and the HO-SVM Wrapper method

Features	
Gesture Volume	Shape Variability
Gesture Area	Bend Error
Ratio of Movement	Length Error
Performance Time	Size Error
Body Pose Density	Speed Error
Body Pose Rate	Acceleration Error
Shape Error	Jerk Error

adults' motions using a statistical test and further validated the subset using a Wrapper method to select the optimal set of features. Overall, the main take-aways, focusing on our optimal subset of features include:

1. Children require more space compared to adults (Gesture Volume, Gesture Area)
2. Children are more inconsistent in how they move their body parts compared to adults (Shape Error, Shape Variability, Bend Error, Bend Variability, Length error)
3. Children move faster compared to adults but show more variations in speed (Performance Time, Speed Error)
4. Children perform motions with less smoothness, as indicated by their higher levels of acceleration and jerk (acceleration error, jerk error).
5. Children performed certain motion types differently from adults, as indicated by the interaction effect between age group and motion type across multiple features. For example, Children and adults were different across 8 out of 13 features for "Swipe across an imaginary screen", "Lift your leg to one side" (6/13), and "Do a forward lunge" (5/13), confirming that the type of motion being performed affects children's motion performance.

6.2 Qualitatively Describing Children's Motion Articulations

Our quantitative results showed that children are generally inconsistent in how they perform motions compared to adults. To better understand in what ways children are inconsistent, thus providing support for our quantitative results, we qualitatively analyzed children's motions. Prior work in the motor development literature has shown that the development of children's motor skills varies by age [23], so we also investigated whether the motion qualities of younger children (6 years) will differ from those of older children during our analysis. To help us with qualitatively analyzing children's motion data, we used an inductive coding approach [87].

For the qualitative analysis, we included some additional motions from the Kinder-Gator dataset: mirror motions (10 motions in total, e.g., Raise your other hand), motions differing in strength (3 motions, e.g., Throw ball), and motions indistinguishable from motions in the representative set of motions using the Kinect v1 alone (2 motions, e.g., Push an imaginary button). These motions had been excluded from our representative set of motions, which we used for our quantitative analysis due to being too similar to other motions. Including these motions in our qualitative analysis will enable us to understand whether children perform motions differently when using dominant vs. non-dominant hand or when the strength required to perform the motion differs. In total, we analyzed 29 motions.

6.2.1 Method

We conducted a qualitative, thematic analysis of the RGB videos of children performing motions (290 motions in total, 29 motions across 10 videos). Three researchers (coder1, coder2, and coder3) participated in the analysis process; coder1 is the author of this dissertation work. Coder1 began with an open coding pass of all 29 motions from two randomly selected videos of children's motions. Then, Coder1 developed a codebook reflecting codes that described objective and subjective qualities of how children perform motions (the full codebook is added to the appendix). The codes were developed based on observable characteristics of children's motions that seemed to differ clearly from adults' motion performance, based on c1's expert understanding about adults' motion performance. These characteristics include how children moved each body part (4 codes, e.g., number of arms [used in] movement), how they moved multiple body parts relative to each other (27 codes, e.g., direction of leg movement), how much effort they used to perform the motions (3 codes, e.g., effort used to move the arms), and the speed of their body movements (1 code: body speed). Then, c1, c2, and c3 coded random subsets from the set of 29 motions in one of the earlier selected videos and met to discuss code disagreements and refine code definitions. Then, coder1 coded different random 10% subsets of the motions (29 motions in total) with coder2 and coder3 respectively, to compute Inter-Rater Reliability (IRR), a measure of the degree of agreement among coders [49], in order to

understand the robustness and reliability of the codebook. We measured IRR using Cohen's Kappa, which ranges from -1 (perfect disagreement) to 1 (perfect agreement) [51].

The coding comparison showed a Kappa value of 0.44, with a 61% average agreement among coder pairs. The Kappa value reflects moderate agreement according to Landis and Koch's scale [73]. Le Bras et al. noted that such moderate agreement "reflects a codebook which was meaningful (objective) but not too restrictive (allowed subjectivity)" [74]. This is especially the case for our codebook as we had four codes (effort_upper_body, effort_arm, effort_leg, and body speed) that required coders to compare a child's movement to that of a referent (i.e., a representative sample of a motion). In addition, we believe that this moderate level of agreement further reflects the variations in how children perform motions, which could have further increased the subjective disagreements among coders. Even for codes that seemed simple and objective (e.g., number of arms bent), we still noticed a high number of disagreements among coders. Children often made movements that were subtle (e.g., slight bending of the arm as a result of the forceful movement of other body parts). Because such movements are not easily noticeable, coders disagreed on whether the movement actually occurred, which affected the objective codes. Finally, coder1, who had the most experience investigating children's motions, coded all 290 motions in the dataset to ensure consistency in the final interpretations. A limitation of this approach is that the codes are not generalizable (i.e., another coder coding the same set of motions may not have the same codes). To mitigate the effect of this limitation on our findings, all three coders met to discuss and agree on the themes that emerged from the analysis.

6.2.2 Results

After coding, using our own expert mental models, we looked through the codes to identify instances in which children performed motions differently from the expected way the motion should be performed (e.g., using two arms to raise a hand instead of one arm). Next, we identified themes that each express a behavior exhibited by the children that caused the instances to occur, thus resulting in variations in children's motion performance. We will now discuss the themes

that resulted from our analysis. While the specific examples we reference within each theme are based on coder1's codes, the themes were agreed upon by all coders of the dataset.

Uncertainty during the motion. Children initially performed movements different from what they apparently intended. We noticed that some children initially started performing a motion with one limb, interrupted the motion of that limb and then started the same motion with the opposite limb. For example, when performing the motion "Raise your arm to one side", (Child A, 6 years old) began raising their left arm. Prior to raising that arm completely to the side, the child dropped the arm and raised their right arm instead. In our dataset, this behavior predominantly occurred in motions involving one arm. Our analysis showed that for such motions, some children moved both arms, indicating that their actions and intentions were not in sync: "Raise your hand" (1 child raised both arms), "Raise your other hand" (2), "Raise your arm to one side" (1), and "Raise your other arm to the other side" (1). We also noticed this behavior more among younger children ($\leq$ 6 years) when making the mirror motions, such as "Raise your other hand" (a mirror of the motion "Raise your hand") and "Raise your other arm to the other side" (3 children in total). Children in this age range are just beginning to differentiate the left and right sides of their body [23] and have shorter attention spans [108], so they will take a longer time processing their thoughts and are more likely to forget what limb they raised previously as compared to older children ($>$ 6 years). Consequently, they change their motion once they realize that it is different from what they intended. Although we only saw this behavior in mirror motions, findings from stroke gesture research suggest that this behavior is common in children's interactions. For example, Shaw et al. [120] found that children have a tendency to trace over gestures they have already drawn (e.g., "A"), indicating that they are changing the gesture to match what they intended.

In addition, children also showed difficulties knowing when to stop a motion. They would maintain a particular pose for a long period of time until the experimenter prompted them to stop. We noticed that this behavior occurred more among younger children. For example, (Child B, 6 years old) raised their arm to one side and maintained the "arm to the side" pose, until the

experimenter prompted them to stop, after which the child then completed the motion. Our analysis further showed that the child completed the motion at a slower speed. We believe this behavior is because younger children sometimes struggle to follow instructions [108].

Uncertainty prior to performing the motion. Children showed confusion when asked to perform some of the motions,indicating that they were unfamiliar with how the motion the motion is to be performed, which led the experimenter to show the motion to the child before the child completed the motion. The confusion occurred among younger children when performing the "Do a forward lunge" motion (2 children showed confusion), a complex movement involving the whole-body. For example, Child F (5 years old) said, "...I do not know" and shook their head when asked to perform this motion. This confusion means that children either knew the motion but did not recognize the name or did not recognize the motion at all. In the case of the former, the experimenter's performance of the motion will help the child recollect how the motion is performed but may also cause the child to modify their motion performance to match that of the experimenter. In the case of the latter, the child is seeing the motion for the first time and will try their best to copy how the experimenter performed the motion. This confusion also resulted in variations in motion performance. Even though the two children were shown the same motion by the same experimenter, they performed the motion differently. For example, (Child C, 6 years old) bent both arms while (Child F, 5 years old) only bent only one arm, as indicated by the "number of arms bent" code. In the US, children of this age (i.e., 5 to 6 years) are either in Preschool or Kindergarten (i.e., just reached the school age). As such, they are still developing their movement vocabulary [30] and will have less experience engaging in formal activities that teach such complex movements (e.g., during Physical Education (PE) classes), which results in confusions regarding how the motion should be performed. For example, in California, PE teachers are not required to teach forward and side lunges until the child reaches grade two (i.e., around 7 to 8 years old) [30].

Differences in motion performance. Even when children were not confused, some children still performed similar motions differently. This variation in motion performance

occurred in motions that required conceptualization of an imaginary object (e.g., “Kick a ball”, “Throw a ball”), motions that required strength (e.g., “Punch”, “Throw a ball as far as you can”) and in the two motions involving the whole-body (“Do a forward lunge”, “Jump”). For example, our analysis showed that three of the ten children threw a ball with both arms while performing the “Throw a ball” motion, two children threw a ball with both arms when performing the “Throw a ball as far as you can”, and two children punched with only one arm. Similarly, for the “Kick a ball as hard as you can” motion, our analysis showed that during the peak pose (i.e., the last pose before children start to return to the default stand pose), the direction of the kick varied. Children’s leg faced one of the following directions: forward (pointing towards the camera; 2 children), down (pointing toward the floor, 2 children), forward-down (pointing in between the camera and the floor, 3 children), and forward-side (pointing forward but not directly at the camera, 3 children). We also noticed that children performed motions differently from the “standard” form in which the motion is most often performed, thus resulting in variations in motion performance. We saw this behavior, for example, during the performance of the “Do a Forward lunge” motion. Even though the standard form of this motion involves bending both legs with one leg forward and one leg back, while the knees are facing the forward direction, our analysis showed that among the eight remaining children who were not confused about the motion (see the previous section regarding uncertainty prior to performing the motion), four children bent one leg, three bent both legs but along varying directions, and one child did not bend any leg. Although variations in children’s motion performance occurred both in younger and older children, we did notice that younger children were more likely to perform motions differently for some of the motions. Younger children ($< 6 years$) accounted for two out of the three children that threw a ball with both arms and for all of the children that threw a ball far with both arms and punched with only one arm.

Furthermore, children were more likely to be inconsistent when performing motions that involved multiple limbs as opposed to simpler motions requiring only one limb (e.g., “Raise your hand”, “Raise your arm to one side”). We noticed that children’s motion performance was

influenced by prior experiences. For example, (Child F, 5 years old) performed the "Throw a ball" motion similar to how a pitcher pitches a ball during a baseball game, suggesting that the child either actively plays or watches baseball. Similarly, a child who actively plays or watches soccer will likely "Throw a ball as far as you can" with both arms to mimic the soccer throw in and "Kick a ball as hard as you can" to mimic a free kick. Further corroborating the role of prior experience, the previous section showed that children with limited experience, in which the experimenter illustrated the motion, will show inconsistencies in motion performance.

Repetitions. Children sometimes performed multiple repetitions of motions that are not typically periodic, thus deviating from the expected form. This behavior was the most prevalent for "Swipe across an imaginary screen" (4 children) and "Swipe across an imaginary screen with the other hand" (4 children), in which the children swiped their arms multiple times even though the expected form involves swiping the hand once either from right to left or left to right. Three other motions were repeated by children unexpectedly as well: "Jump" (2 children), "Jump as high as you can" (1) and "Punch" (1). This behavior of performing multiple repetitions was prevalent in younger children; only one older child (Child G, 8 years old) swiped multiple times when performing the "Swipe across an imaginary screen" motion. Gesell [41] noted that children at ages 5 to 6 years old are still very active,boisterous, and love to engage in play that involves gross motor activity, such as swinging and jumping, suggesting that this repetitive behavior may be because of children's increased enthusiasm when performing the motions. Our observations of children's overall attitude corroborate this idea as we saw that they were often excited and made random remarks that expressed their enthusiasm when performing motions. For example, (Child A, 6 years old) laughed and clapped in excitement after performing some of the motions. Our analysis corroborated this idea as it showed that whenever this repetitive behavior occurred, children performed the motions at fast speeds and put in extra effort in the key body parts required to perform the motions (e.g., arms when swiping and legs when jumping).

Unrelated movements. Children also sometimes made movements irrelevant to the actual motion being performed. Our analysis indicated that children moved and bent unnecessary body

parts because of unrelated extra movements. For example, (Child F, 5 years old) moved two arms during the motion "Motion someone to stop," with the additional arm moving due to the child swaying both arms forward and backward, a movement unrelated to the motion being performed. Likewise, (Child D, 5 years old) continuously swayed their body throughout their performance of the "Point at the camera" motion resulting in lower body movement even though this upper-body motion only requires the movement of one arm. For lower-body motions, unrelated movement sometimes meant movement of upper body parts that are not required to perform the motion. For example, (Child D, 5 years old) placed both arms on their hips toward the end of the "Lift your leg to one side" motion, thereby adding unrelated arm movements to this lower-body motion.

We also noticed that these unrelated movements often occurred toward the end of the motion. For example, (Child F, 5 years old) always placed both arms on their hips instead of placing their arms at the resting position after completing each motion. The same child also continuously swung their arms on their way back up during the "Touch your toes" motion. We also noticed that these unrelated movements were sometimes caused by children getting distracted. For example, (Child G, 8 years old) touched their hair while performing the "Raise your hand" motion while (Child F, 5 years old) stretched their upper body in an unrelated movement while performing the "Raise your arm to one side" motion. Although this behavior sometimes occurred in older children, it was more prevalent in younger children, likely because they have shorter attention spans and struggle to follow instructions [108].

Unintentional movements. Children also sometimes moved extra limbs unintentionally when performing motions. We refer to a body part as unintentionally moving if the user is actively moving the body part but without the user being fully aware of their action. In contrast to unrelated movements, unintentional movements are relevant to the actual motion being performed (e.g., to aid in the completion of the motion) but are not required to perform the motion. Our analysis showed that this behavior often occurred whenever children were performing motions involving a shift in their center of gravity, causing children to move extra body parts not required to perform the motion. Hence, this behavior was predominant in lower-body motions that

involved lifting the feet off the ground (e.g., “Lift your leg to one side, “Bend your knee”). Although lower-body motions do not require movement of upper body parts, children often moved their arms apparently unintentionally during the performance of these motions: “Lift your leg to one side” (9 of 10 children moved their arms), “Lift your leg to the other side” (10), “Bend your knee” (5), “Bend your other knee” (4), “Kick a ball” (10), “Kick a ball as hard as you can” (10), ‘Kick a ball with your other leg” (10), “Kick a ball as hard as you can with your other leg” (10).

Our analysis also showed that children tended to move extra limbs unintentionally when they overestimated the height or intensity needed to perform the motion, as indicated by the effort they used to make the movement and their ability to maintain the peak pose (i.e., the pose that captures the actual movement being performed). For example, during the “Lift your leg to one side” motion, (Child E, 8 years old) lifted their leg so high but overestimated how high their leg could reach. As such, the child was unable to maintain the peak pose and faltered (i.e., the leg dropped to the ground), thus interrupting the motion. Subsequently, when the child raised their leg that high again, they simultaneously raised their arms. This phenomenon was exhibited by five of the ten children who performed this motion, of which four children put in extra effort into moving their legs. Three children faltered when performing the “Kick a ball with your other leg” motion, of which two children put extra effort into moving their legs. Our analysis further showed that children are likely to falter when they raise their non-dominant leg to the side. For the “Lift your leg to one side” motion, both children who raised their non-dominant leg to perform the motion, faltered while for the “Lift your leg to the other side” motion, four of the six children who raised their non-dominant leg to perform the motion, faltered. However, we did not notice this phenomenon in the other lower-body motions.

Although not as prevalent as in lower-limb motions, our analysis showed that children also moved their lower-body parts unintentionally when performing upper-body motions wherein the children’s center of gravity can shift depending on how the motion is performed. That is, motions involving putting hands on the hips and leaning to side and motions involving throwing a ball:

"Put your hands on your hips and lean to one side" (8 of 10 children moved their legs to perform the motion), "Put your hands on your hips and lean to the other side" (9), "Throw a ball" (8), "Throw a ball as far as you can" (10), "Throw a ball with your other leg" (7), and "Throw a ball as far as you can with your other arm" (10).

We believe that children unintentionally move extra limbs to perform these motions because they are still developing their motor and postural abilities [23, 55] and have less experience coordinating their limbs to perform movement compared to adults, resulting in motions that are not well-coordinated. Hence, they move extra body parts unintentionally to maintain balance when performing motions. This unintentional movement of extra limbs will lead to variations in how children perform motions. For example, our analysis showed that when performing the "Kick a ball" motion, three children moved their right arm along the XY direction to get to the peak pose, five children moved along the YZ direction, and the remaining two children moved along the XZ direction. Similarly, during the peak pose, six children had their right arm facing the side-down (i.e., perpendicular to the body but facing toward the floor) direction, three children had their right arm facing backward (away from the camera), while the last child had their right arm facing down.

Although most of the unintentional movements resulted from shifts in children's center of gravity as they performed motions, our analysis also showed that some unintentional movements of extra limbs resulted as a consequence of the intensity with which children moved other body parts. For example, when performing the "Punch" motion, (Child B, 6 years old)'s left arm moved due to the intensity with which the child punched with their right arm. Unintentional movements arising from this type of situation are often subtle, which makes them difficult to notice with the naked eye but can still be accurately tracked by the motion sensor.

6.3 Discussion

So far, in this Chapter, we have focused on understanding the inconsistencies in how children and adults perform motions. To do this, we identified features that quantitatively described users' motion performance and evaluated these features on a subset of children's and

adults' motions to reveal differences. Our results showed that children move differently from adults in ways that can be quantified with specific posture- and joint-motion-based articulation features. For example, children move faster and are more inconsistent in how they move their body parts to perform motions, as compared to adults. Furthermore, we qualitatively analyzed children's motions to understand in what ways children are inconsistent when performing motions. In this section, we looked across the features to identify themes, in which a theme is an inference from the result (e.g., children require more space compared to adults). We grouped these themes to identify dimensions along which children's motions differ from adults' motions, indicating children's motion qualities, and use findings from prior work and our qualitative analysis to triangulate our quantitative findings. We also discuss the practical significance of our articulation features in this section.

6.3.1 Children's Natural Motion Qualities

The dimensions along which children's motions differ from adults' motions include:

Speed: This dimension relates to how fast children perform motions. Children move faster than adults. Children are consistently faster than adults when performing motions (based on features like performance time and time error). This finding echoes and deepens prior work, which showed that children move faster than adults for walking and running motions [4]. Our results also show that children move with higher speeds but exhibit more variation in the speeds they use to perform motions (average gesture speed, speed error). Since speed measures the rate of joint movement and time is consistent, these variations mean that children are inconsistent in how they move their body parts to perform motions. Findings from our qualitative analysis corroborate this idea as it showed that when children move at higher speeds, they move extra limbs unintentionally, resulting in variations in the body parts children move to perform motions. Prior work in stroke gesture research has shown that gestures articulated at faster speeds have higher inconsistencies compared to gestures articulated at slower speeds [135]. Hence, there is a speed-accuracy trade-off relating to children's motion performance; they move quickly but are

more inconsistent (i.e., less accurate) in how they perform motions. This trade-off emerges because children are still developing their motor abilities [23].

Intensity: This dimension relates to the amount of effort used to perform motions. Children perform more exaggerated motions, therefore more intense, motions as compared to adults. Children and adults differ in the appearance of their motions when space and time are considered. Children's postures are less dense and are completed faster than adults (Body pose density, Body pose rate) since they require more space and move faster (Gesture volume, Performance time). Hence, children's motions will appear more exaggerated as compared to adults, thus requiring more effort [15, 18]. Prior work in biomechanics supports this finding as it noted that exaggerated postures require more energy [119]. In addition, prior work in exercise motions also noted that exaggerated motions require less time [7]; an assertion evidenced by our Performance Time and Body pose rate findings. Our qualitative findings further support our quantitative results as it showed that children exaggerate motions. They perform repetitions of postures that are non-periodic (e.g., jumping twice) and whenever such repetitions occur, children always put in extra effort to move key body parts (e.g., extra effort in their legs to jump). Like speed, intensity also plays a role in the inconsistencies children show when performing motions. Since children are still developing their motor skills [23] and move fast, the additional effort they use to perform motions could result in loss of control over body parts during movements, resulting in loss of balance (a behavior that leads to jerky motions).

Smoothness: This dimension relates to how well children move each body part that is necessary to performing a motion. Children's motions are less smooth as compared to adults' motions. Children jerk inconsistently when performing motions (jerk error) but move with higher levels of jerk in their motions (average gesture speed, performance time). Therefore, children are more likely than adults to make jerky motions, which explains why children are inconsistent in how they move body parts (shape error). Furthermore, the inconsistency in the uniformity of their shape errors means that children are also inconsistent in the ways in which they are inconsistent (shape variability). Prior work has noted that children in the age range we considered (i.e., 5 to 9)

are still developing their motor abilities [23]. Therefore, we believe that they perform motions less smoothly as compared to adults because they have less expertise controlling body parts to perform motions. Our qualitative findings further supports this idea as it showed that children find it difficult to control their body parts when performing motions in which they have to maintain their balance (e.g., "Throw a ball as far as you can"). As mentioned in the previous section, speed and intensity also play a role in the smoothness of motions. Children are more likely to lose control over body parts when the motion is performed with high speed and forceful intensity, especially for complex motions involving the whole-body ("Do a forward lunge" and "Jump"). Our Shape Error finding validates this idea as it showed that children are not as consistent as adults when performing both motions.

Coordination: This dimension relates to how well children move body parts relative to each other. This quality is closely related to smoothness, but coordination involves multiple body parts. Children make less well-coordinated multi-limb movements as compared to adults. Motions often require a lateral shift of an individual's center of gravity and body balance due to postural changes [119]. Motions involving movements of lower body parts (e.g., "Kick a ball as hard as you can", "Lift your leg to one side", and "Bend your Knee") shift a user's center of gravity once they lift their foot. We know from our other themes that children move with high speeds and forceful intensity, behaviors that result in a shift in body balance [119]. However, children are still developing their postural stability [55], indicating that they will often overestimate the speed and intensity at which a motion should be performed and falter due to loss of balance. Prior work has noted that children move their arms when performing lower body movements to maintain balance [55], which may explain why children move their upper body more than their lower body when performing movements, as compared to adults (Ratio of Movement). Therefore, we see that children move extra body parts that are not necessary to motion performance, in order to stabilize their body once they begin to falter. Findings from our qualitative analysis support this idea as it showed that children overestimate the effort they need to perform motions involving shifts in an individual's center of gravity (e.g., "Lift your leg to one

side"), so they falter. Subsequently, they move their arms to maintain balance. Prior work further noted that how users account for a shift in body balance will determine their degree of coordination [55], with well-coordinated movements requiring control over body parts. Since children move these extra body parts in response to a loss in balance, they will not move these extra body parts as intentionally (i.e., with full awareness) as the key body parts they are actively moving. Therefore, children will have little to no conscious control over how these body parts move (e.g., the direction of movement). Combined with the smoothness dimension, the above statements indicate that children not only find it difficult controlling one body part in isolation, but will also find it difficult coordinating multiple body parts.

Prior work further noted that the movement of extra body parts results in inconsistencies in how children perform motions [5]. The authors computed the degree of agreement among the actively moving joints when performing lower body motions [5]. The degree of agreement is defined as the total number of unique joint combinations used to perform a motion. They found that children had a lower degree of agreement, indicating higher inconsistency, for lower-limb motions due to some children moving upper body parts (e.g., arms) even though such body parts were not necessary for motion performance.

6.3.2 Practical Significance of Articulation Features

Our feature validation step provided a set of articulation features that can differentiate child from adult motion with about 81% accuracy using an SVM classifier. This finding has important implications for designing emerging technologies in the real-world. For example, smart home personal assistants currently use voice, a natural communication modality, for interaction. To fulfill Mark Weiser's vision that technologies will be interconnected and integrate seamlessly with the environment [144], future trends of this technology (i.e., smart environments) will likely benefit from including motions, another natural communication modality, for interaction and authentication [95]. In our past work, we published the MMGatorAuth dataset to facilitate active and passive interaction in smart environments using voice and motions [95]. Approximately 32% of US households already have a smart home device and it is expected to grow to 57% by 2025

[136]. Since households sometimes comprise both children and adults, it becomes important that the device can accurately distinguish children's motions from adults' motions to mitigate security and usability concerns. We provide an initial set of features that can be used to ensure accurate recognition of children's motions. Future work can identify new child-specific features based on children's natural motion qualities we identified to improve the recognition performance.

CHAPTER 7
DESIGN IMPLICATIONS

As a reminder, the goal of this dissertation work is to investigate the differences between children's and adults' motions to inform an understanding of how children perform motions (i.e., their natural motion qualities). We found that children's natural motions differ from adults' motions along four dimensions: speed, intensity, smoothness, and coordination. Children perform motions that are quantifiably faster, more intense, less smooth, and less coordinated. We also found that the type of the motion also plays a role in the variations children show when performing motions. Based on our findings, we now propose guidelines for improving motion recognition algorithms and designing motion applications for children.

7.1 Motion-Based Application Design

Motion-based applications (e.g., exertion games [141, 80]) are becoming increasingly popular among children as researchers and practitioners are using these applications to target children's needs (e.g., increasing the time children spend engaged in physical activity). We propose two guidelines for designing motion-based applications based on findings from our dissertation work, that can help improve children's interactive experiences when using these applications:

Favor simpler motions. Children are less consistent when performing complex movements that involve the whole-body. Therefore, designers of motion applications should favor simple motions that children are familiar with (e.g., upper-body motions requiring a single limb) instead of more complex whole-body motions (e.g., lunges). However, favoring simple motions may not always be feasible (e.g., exercise games require some complex exercise motions to ensure children achieve moderate-to-vigorous physical activity [80, 141]). Designers should provide opportunities for children to practice complex motions to get them more familiar with how the motion is performed.

Be flexible about space requirements. Children require more space to perform motions and are inconsistent in the space they require to move body parts. Therefore, designers of whole-body applications should customize space requirements to ensure that children have enough space (with respect to area and volume). Furthermore, designers should program the

application to verify proper space allowances (e.g., using the depth camera to detect hazards) to prepare for scenarios in which children make exaggerated motions that will require even more space. Space verification is important to prevent injury [45].

7.2 Motion Recognition Systems

Motion-based applications often require accurate recognition of motions to ensure meaningful interactive experiences. Even though these applications are becoming increasingly popular among children, motion recognition systems are usually trained on adults' motions [22] even though my past work has shown that children move differently from adults [64]. In this dissertation work, we investigated how and why children move differently from adults. Based on our findings, we now propose a set of design guidelines that future research can adopt to help tailor motion recognition systems to children's motion qualities for accurate recognition. The guidelines proposed from our findings have the potential to improve the performance of motion recognition systems on children's motions since prior work has found several correlations between global-level features and the performance of motion recognition systems on adults' motions [132] and 2D stroke gesture research has also shown a correlation between articulation features and children's stroke gesture recognition performance [120].

Favor looser pointing approaches. In Chapter 4, we showed that how we adapted template-based stroke gesture recognizers that use one-to-one matching to motions [5]. However, children make jerky motions and are inconsistent in how they move their body parts to perform motions (shape error, shape variability). Therefore, we recommend designers of motion recognition systems for children to use less stringent point-based approaches. For example, designers can use something similar to the many-to-one approach used in the $P+ stroke gesture recognizer [133]. This approach accounts for variations along an articulation path by choosing the best point in a template path that matches a given point in the candidate path, such that multiple points in the template path can be assigned to one point in the candidate path during recognition. This approach will be less affected by shape errors and shape variabilities, thus improving recognition.

Expect repetitions for non-periodic motions. Children sometimes repeated motions that could or should have only been performed once, such as Jumping or Swiping, which relates to their tendency to exaggerate motions. Therefore, we recommend that designers of motion recognition systems should account for an unpredictable number of repetitions during recognition.

Account for extra limb movements. We saw children often moved extra limbs to perform motions, which resulted in inconsistencies in their motion performance. Therefore, designers of motion recognition systems should consider only the joints that users are actively moving intentionally. Prior work in motion recognition with the filterJoint method found an increase in recognition accuracy when only the minimum sufficient subset of the joints tracked by the motion sensor is considered [5]. For example, designers can use the filterJoint method to select all actively moving joints and use only these joints during recognition to improve accuracy.

Consider directional differences. In Chapter 6, we noted that children sometimes get confused differentiating their left from their right body parts. Furthermore, children have little to no conscious control over the direction of movement of upper body parts when performing motions requiring them to maintain balance. Therefore, designers of motion recognition systems should always apply normalization steps that can remove confusions from motion instances of the same type due to directional differences. For example, designers can apply the translation and rotation normalization steps (e.g., those used in our adaptation of template-based stroke gesture recognizers [5]) to pre-process joint paths before recognition.

Expect differences in motion performance based on age.[1] Findings from our qualitative analysis show that younger children ($<$ 6 years) perform motions differently from older children, due to factors such as limited experience performing motions. Therefore, designers of motion recognition systems should consider accounting for age ranges during recognition. For example, designers can consider including weights during the matching process of template-based recognizers such that children's motions are more likely to be matched to motions from children

[1] More research is needed to fully understand the effect of age on the variations in how children perform motions.

with similar age groupings (i.e., matching younger children to younger children and older children to older children).

Minimize inconsistencies during the collection of motion sets. In Chapter 6, our qualitative analysis showed that children are inconsistent in their motion performance due to their unfamiliarity with the motion to be performed. Therefore, we recommend that when researchers collect motion sets, they should design their study such that each motion is performed at least twice, with the first motion being practice. By doing this, we expect that children would have gotten more familiar with the motion to be performed during the second trial, thus minimizing any potential confusion. Prior research in 2D touchscreen devices has used this approach of excluding the first trial as a practice to prevent errors arising from children being unfamiliar with the task performed [150].

CHAPTER 8
CONTRIBUTIONS AND FUTURE WORK

Here we detail the contributions of this dissertation work to the field of human-computer interaction and present opportunities for future research to build upon the contributions of this dissertation.

8.1 Identifying Features That Quantitatively Describe Motions

In addition to Vatavu's [132] global-level features, we identified a set of joint-level features that quantitatively describes the properties of joint articulation paths. These joint-level features were inspired by prior work on relative accuracy features in stroke gesture research [135], features for tracking mouse paths [56], and features from Laban Movement Analysis [72]. Our analysis of children's and adults' motions using these features showed that they were useful in quantifying the differences between their motions.

Future research can use our set of features to investigate the differences between motions for other populations. For example, future research can use these features to investigate the differences between typically developing children and children with disabilities whose motor skills may not be as developed (e.g., children experiencing cerebral palsy). Future research can also use these features to differentiate between the motions of older adults and younger adults. Older adults (>65 years) experience a decline in their motion abilities [117], so they are likely to perform motions differently than younger adults. In addition, since our features were effective in characterizing children's natural motion qualities, future research can use these features to evaluate whether the motion qualities of children's motions generated using style translation techniques are indicative of the motion qualities of actual children. For example, Dong et al. used dynamic scaling laws to translate adults' motions to children's motions [31]. The effectiveness of the scaling technique can be evaluated by investigating the similarities between the motion qualities of the style-translated child and an actual child.

8.2 Understanding how Children Perform Motions

Throughout this dissertation work, we conducted multiple studies and analysis to understand how children move differently from adults, from which we identified four dimensions: speed, intensity, smoothness, and coordination. We found that children's natural motions are

performed at faster speeds and with higher intensities and are less smooth and less coordinated, than adults' motions.

A limitation of our work is that we only considered between-user consistency, which means that we don't know how consistent each child is across multiple instances of their motions. Future research can consider using our articulation features to analyze within-user consistency in one of two ways: (1) periodic motions, in which the same postures occur multiple times over a brief time period, such as multiple jumping jacks; or (2) multiple separate repetitions of the same motion type, at different time points. While Kinder-Gator [3] does include some periodic motions which we did not analyze here, to the best of our knowledge, there is as yet no dataset of children's motions that include multiple separate repetitions.

In addition, our dissertation work only presents initial observations regarding how age impacts motion performance. Future research will need to quantitatively analyze children's motions to better understand the differences between younger and older children's motion performance for more conclusive findings. However, there is currently no available dataset of children's motions within the age range we considered to enable such an analysis. Future research can collect more examples of children performing natural motions, to better understand in what ways factors, such as age, gender, and grade level, affect children's motion performance.

Another limitation of our work is that our motion data was tracked using the Kinect v1 sensor [89], which was the only available low-cost sensor for tracking whole-body motions at the time of data collection. However, the Kinect v1 is very prone to tracking issues. There are several factors that contribute to inaccurate tracking of the Kinect. For example, prior work has noted that the Kinect v1 has less accuracy tracking joints in the lower body, especially the hip joints and occluded joints [140]. The Kinect v1 also has a narrower field of view [129], as compared to sensors, such as the Vicon (Vicon, Oxford, UK), so the user's height, the placement of the Kinect, and the users' distance from the Kinect, are all factors that will affect the tracking accuracy (e.g., the accuracy reduces the farther away the user is from the Kinect [48]). Therefore, the Kinect v1 has a high signal-to-noise ratio, which reduces the quality of the motion data due to the noise

introduced. Although we did account for noise in the data from tracking issues by smoothing the joint path using an exponential smoothing filter and designing a filterJoint method that can eliminate joints that move due to tracking issues, we recommend that future research that wants to investigate users' motions should use sensors with better tracking accuracy. For example, the Vicon sensor tracks users' motions with high enough accuracy that it is often regarded in the research field as the gold standard [2]. However, the Vicon sensor is not always practical to use because it is expensive, not portable, and requires the placement of markers on the body, making it intrusive [79]. Future research in lieu of the Vicon, can use low-cost, markerless sensors with higher tracking accuracy than the Kinect v1. For example, the most recent version of the Kinect, known as the Azure Kinect DK [90]), has been shown in the gait literature to achieve higher tracking accuracy compared to previous versions of the Kinect [2, 129].

8.3 Improving Recognition Rates for Children

Based on the new insights we learned about children's motion performance, we proposed design guidelines for improving motion recognition systems for children (Chapter 7). Future research can adopt these guidelines to tailor motion recognition systems to children's motion qualities and evaluate the performance of the system to investigate whether there is an improvement in performance on children's motions. If there is an improvement, future research can further investigate whether there is a correlation between improved recognition rates and children's interactive experiences in motion-based applications. Prior work has found that the precision of motion recognition systems is associated with increased immersion in exertion games [98]. Furthermore, our work shows that algorithms trained on adults' motions perform poorly on children's motions due to the differences that exist between children and adults. This finding has important implications on research for other modalities, such as facial expressions and speech. Similar to motions, recognition systems for these modalities are typically trained on adults. We know from prior work that children are different from adults with respect to their neuromuscular, cognitive, and motor abilities [23, 114], thus suggesting that algorithms trained on adults will likely perform poorly on input from children. Future research can leverage the systematic analysis

used in this work to understand the differences between children and adults with respect to the specific modality being considered (e.g., differences in their speech) and subsequently, tailor these recognition algorithms to children qualities. We also use this opportunity to make a call to action for future research to consider including populations different from the typically developing adult population (e.g., children, older adults) during the design of recognition algorithms.

However, it is important to note that we only looked at three types of motion recognition algorithms in this dissertation work, two of which were template-based algorithms and the last one being a traditional machine-learning algorithm. Advancements in machine learning have resulted in the development of algorithms that use neural networks for the automatic recognition of motions. These algorithms have been shown to achieve higher accuracy compared to traditional motion recognition algorithms on adults' motions (see section 2.3.2.2), thus having the potential to perform better on children's motions. However, neural networks require large numbers of samples of data for training [77], meaning that it is not currently feasible for recognition of children's motions due to the lack of publicly available child motion datasets. To use neural networks, future research will need more data from children. However, because recruiting children for motion experiments can be challenging due to their shorter attention spans [108], collecting real samples of children performing motions can prove difficult. Researchers can instead consider generating synthetic data by artificially modeling children's motions, for example, using style translation [31] and and motion synthesis [156, 82]. In motion synthesis research, extensive work has focused on generating human motions from existing data in motion datasets using approaches that rely on autoregressive models, statistical learning, and deep learning [156]. To ensure that synthetically generated motions are similar to that of an actual child performing motions, researchers can use the articulation features presented in this dissertation work to compare both motions.

CHAPTER 9
CONCLUSION

Motion-based applications usually include motion recognition algorithms that can accurately recognize the specific sets of motions that the applications support (i.e., motion sets). Accurate recognition of motion sets plays an important role in users' interactive experiences. However, skeleton-based recognition systems, which accept as input the positions of joints tracked by a motion sensor in 3D space over time, perform poorly on children's motions compared to adults' motions (Chapter 4). In this dissertation work, we focused on understanding how children perform motions to tailor motion recognition systems to children's motion qualities for accurate recognition of their motions. Specifically, we answered the following research questions (RQ) in this dissertation:

RQ1. What are the differences between children's and adults' motions?

RQ2. What inferences can we make from children's and adults' motion differences to help tailor motion recognition systems to children's motions?

9.1 Research Question 1- Differentiating Children's and Adults' Motions

To answer this research question, we focused on understanding the nuances in how users articulate motions, which we achieved by identifying the joints that are critical to performing motions and identifying features that can quantify how the motion is produced to aid in the analysis of child and adult motion.

9.1.1 Analyzing Joints That are Critical to Motion Performance

We designed a method that can facilitate the investigation of variations in how users move body parts as a perform motions (Chapter 5). Our method, which we termed filterJoint, filters out noisy or unimportant joint motion paths using standard deviation and K-means clustering [53] iteratively to select the set of joints that are actively moving during a motion. We evaluated our method on a subset of adults motions, using an adaptation of a template-based stroke gesture recognizer, we developed and found that our method (90.7% [SD = 6.8%) outperformed a baseline method including all joints (81.4% [SD = 6.9%]). We computed the degree of agreement, which we defined as the total number of unique joint combinations selected within a motion type. Our analysis of children's and adults' motions using our filterJoint method showed

that children had a lower degree of agreement compared to adults, especially for complex movements involving increased coordination among many joints to perform the motion (e.g., Jump). Therefore, children are more inconsistent when performing motions compared to adults.

9.1.2 Identifying Features That Quantitatively Describe Users' Motion Performance

Having established that children are more inconsistent in how they perform motions compared to adults, we sought to understand how and why children are more inconsistent in their motion performance. To do this, we quantified the differences between children's and adults' motions. Inspired by prior work in stroke gesture research [135], we focused on identifying features that quantitatively describe how users perform motions. We begin by analyzing children's and adults' walking and running motions using gait features and found that children move faster and with more energy compared to adults (Chapter 5). However, because gait features are only applicable to periodic motions, in which the same postures occur multiple times over a brief time period, we sought to identify features that are generalizable to a broad set of motions. To do this, we defined a set of 24 articulation features (11 of which are new) that captured properties related to the themes (e.g., length, time, and effort) (Chapter 6). Our articulation features comprised 13 features from prior work [132] that describe motions globally based on the overall posture of the body (global-level features). We also defined a set of 11 new features that characterize properties of a joint moving in 3D space (joint-level). Analysis of these features on a subset of children's and adults' motions showed that children move differently from adults in ways that can be quantified with specific posture- and joint-motion-based articulation features. To support results from our quantitative work, we also qualitatively analyzed children's motions to understand how children move differently from adults.

9.2 Research Question 2-Generating Design Guidelines

We looked across the features to identify themes, in which a theme is an inference from the result (e.g., children require more space compared to adults). From grouping these themes, we identified four dimensions along which children's motions differed from adults motions, namely speed, intensity, smoothness, and coordination (Chapter 6). Children's natural motions are less

smooth and less coordinated, and are performed at faster speeds with higher intensities, than adults' motions. Based on our findings, we proposed guidelines for designing motion recognition systems and motion applications for children (Chapter 7). We hope that designers and researchers can adopt these guidelines to tailor motion recognition systems to children's motion qualities for more accurate recognition and improve children's interactive experiences in motion-based applications.

9.3 Contributions

The work presented in this dissertation offers several contributions to the field of Human-Computer Interaction, as detailed in Chapter 8. First, we established a set of features that quantify the differences between children's and adults' motions. We also informed an understanding of how children perform motions by identifying the dimensions along which children move differently from adults. We presented guidelines for tailoring motion-based applications to children's motion qualities to accurately recognize children's motions.

APPENDIX A
PUBLICATIONS

A.1 Journal Articles

- Eakta Jain, Lisa Anthony, Aishat Aloba, Amanda Castonguay, Isabella Cuba, Alex Shaw, and Julia Woodward. 2016. Is the Motion of a Child Perceivably Different from the Motion of an Adult?. ACM Trans. Appl. Percept. 13, 4, Article 22 (July 2016), 17 pages. DOI: http://dx.doi.org/10.1145/2947616.

A.2 Conference Papers

- Aishat Aloba, Lisa Anthony. Characterizing Children's Motion Qualities: Implications for the Design of Motion Applications for Children. 2021. International Conference on Multimodal Interaction (ICMI '21): 8 pp.
- Aishat Aloba, Sarah Morrison-Smith, Aaliyah Richlen, Kimberly Suarez, Yupeng Chen, Damon Woodard, Jaime Ruiz, and Lisa Anthony. Multimodal User Authentication: Survey of User Attitudes. 2021. 18 pages. [In preparation].
- Aishat Aloba, Gianne Flores, Jaida Langham, Zari McFadden, Julia Woodward, Lisa Anthony. Strategies for Motivating Children in Physical Education Classrooms: Implications for Design of Exertion Games. 10 pages. [In preparation].
- Aishat Aloba, Julia Woodward, and Lisa Anthony. FilterJoint: Toward Understanding Whole-Body Gesture Articulation. 2020. International Conference on Multimodal Interaction (ICMI '20): 213–221.
- Sarah Morrison-Smith, Aishat Aloba, Hangwei Lu, Brett Benda, Shaghayegh Esmaeili, Gianne Flores, Jesse Smith, Nikita Soni, Isaac Wang, Rejin Joy, Damon L. Woodard, Jaime Ruiz, and Lisa Anthony. 2020. MMGatorAuth: A Novel Multimodal Dataset for Authentication Interactions in Gesture and Voice. International Conference on Multimodal Interaction (ICMI '20): 370–377.
- Ziyang Chen, Yu-Peng Chen, Alex Shaw, Aishat Aloba, Pavlo Antonenko, Jaime Ruiz, and Lisa Anthony. Examining the Link between Children's Cognitive Development and Touchscreen Interaction Patterns. 2020. International Conference on Multimodal Interaction (ICMI '20): 635–639.
- Aishat Aloba, Annie Luc, Julia Woodward, Yuzhu Dong, Eakta Jain, and Lisa Anthony. 2019. Quantifying Differences between Child and Adult Motion based on Gait Features. International Conference on Human-Computer Interaction (HCII '19). Springer, 385-402. [Invited Paper].[1]
- Aishat Aloba, Gianne Flores, Julia Woodward, Alex Shaw, Amanda Castonguay, Isabella Cuba, Yuzhu Dong, Eakta Jain, and Lisa Anthony. 2018. Kinder-Gator: The UF Kinect Database of Child and Adult Motion. In Proceedings of Eurographics '18: 4 pages.

[1]This publication was not refereed.

- Soni Nikita, Aishat Aloba, Kristen Morga, Pamela Wisniewski, and Lisa Anthony. 2019. A Framework of Touchscreen Interaction Design Recommendations for Children (TIDRC): Characterizing the Gap between Research Evidence and Design Practice. Proceedings of the Conference on Interaction Design and Children (IDC '19): 419 – 431. [Refereed].

- Julia Woodward, Alex Shaw, Aishat Aloba, Ayushi Jain, Jaime Ruiz, and Lisa Anthony. 2017. Tablets, Tabletops, and Smartphones: Cross-platform Comparisons of Children's 125 Touchscreen Interactions. In Proceedings of the International Conference on Multimodal Interaction (ICMI '17): 5-14.

- Yuzhu Dong, Sachin Paryani, Neha Rana, Aishat Aloba, Lisa Anthony, and Eakta Jain. 2017. Adult2Child: Dynamic Scaling Laws to Create ChildLike Motion. In Proceedings of MiG '17, Barcelona, Spain, November 8–10, 2017, 11 pages.

A.3 Refereed Conference Posters and Doctoral Consortium

- Aishat Aloba, Katarina Jurczyk, Julia Woodward, Lisa Anthony. Identifying Cues that make Children's Motions Perceivably Different from Adults' Motions. 2021. International Conference on Multimodal Interaction (ICMI '21): 4 pp. [In review].

- Aishat Aloba. Tailoring Motion Recognition Systems to Children's Motions. 2019. International Conference on Multimodal Interaction Doctoral Consortium (ICMI '19). 457 – 462.

- Aishat Aloba, Gianne Flores, Jaida Langham, Zari McFadden, John Bell, Nikita Dagar, Shaghayegh Esmaeili, and Lisa Anthony. 2020. Toward Exploratory Design with Stakeholders for Understanding Exergame Design. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '20). To appear.

- Aishat Aloba, Gabriel Coleman, Triton Ong, Shan Yan, Dehlia Albrecht, Marko Suvajdzic, and Lisa Anthony. 2017. From Board Game to Digital Game: Designing a Mobile Game for Children to Learn About Invasive Species. In Proceedings of ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '17): 375-382.

- Yuzhu Dong, Aishat Aloba, Lisa Anthony, and Eakta Jain. 2018. Style Translation to Create Child-Like Motion. In Proceedings of Eurographics'18: 2 pages.

APPENDIX B
KINDER-GATOR DATASET

Table B-1. Motions performed by participants in the Kinder-Gator dataset for which the motion was demonstrated to the participant by a researcher during the collection of the dataset

ID	Agegroup	Motions
106	Child	Forward lunge, Put your hands on your hips and lean to the side
290	Child	Raise your arm to one side, Bend your knee, Swipe across an imaginary screen in front of you, Fly like a bird, Point at the camera, Forward lunge
337	Child	Bend your knee, Lift your leg to one side, Make the letter (M, K) with your body
342	Child	Make the letter (A, K) with your body
888	Adult	Forward lunge
921	Adult	Forward lunge, Make the letter M with your body

APPENDIX C
KINDER-GATOR DATASET: ADULT DEMOGRAPHIC QUESTIONNAIRE

Participant #: _____

1. Sex [Male][Female]
2. Age _____years
3. Level of Education Finished [High school][Some college][Undergraduate degree][Graduate degree][Other:]
4. Handedness [Left][Right][Both]

Please tell me about which devices you own or have in your house:

	I own one	My family owns one	I've never heard of this
XBox Kinect			
Nintendo Wii or Wii U			
Sony PlayStation EyeToy			

Please tell me about your use of the following types of devices:

	Tried it once or twice	I use it sometimes	I use it daily or often	I have never used one	I have never heard of it
XBox Kinect					
Nintendo Wii or Wii U					
Sony PlayStation EyeToy					

APPENDIX D
KINDER-GATOR DATASET: CHILD DEMOGRAPHIC QUESTIONNAIRE

Participant #: _____

1. Sex [Male][Female]
2. Age _____years
3. Highest Grade Level Completed _____
4. Handedness [Left][Right][Both]

Please tell me about which devices you own or have in your house:

	I own one	My family owns one	I've never heard of this
XBox Kinect			
Nintendo Wii or Wii U			
Sony PlayStation EyeToy			

Please tell me about your use of the following types of devices:

	Tried it once or twice	I use it sometimes	I use it daily or often	I have never used one	I have never heard of it
XBox Kinect					
Nintendo Wii or Wii U					
Sony PlayStation EyeToy					

APPENDIX E
FILTERJOINT METHOD: PSEUDOCODE

Step 1: Apply the filterJoint method to both the test motion C and all the motions in the training set T, where each motion has N 3D joints. Note: This method is only called if using the filterJoint method. When using the baseline method, skip this step. The filterJoint method relies on a threshold, thresh which is the average absolute difference between the centroids of the clusters of all motions in the dataset (i.e., the THRESHOLD method is only called once).

FILTERJOINT(C, T)

```
SD ← COMPUTE-STANDARD-DEVIATION(C)
A ← GET-ACTIVE-JOINTS(SD)
for i from 0 to N do
    if i not in A then
        REMOVE(C, j^i)
        REMOVE(t, j^i) foreach motion t in T
return <C,T>
```

COMPUTE-STANDARD-DEVIATION(A)

```
SD ← empty-list
foreach joint j^i
    for i ≥ 0 in C do
        j^i ← EXPONENTIAL-MOVING-AVERAGE(j^i)
        S ← STANDARD-DEVIATION(j^i)
        APPEND(SD, S)
return <SD>
```

GET-ACTIVE-JOINTS(A)

```
active_joints ← empty-list
inactive_joints ← empty-list
<centroid_1, centroid_2, labels, index> ← KMEANS(A)
    for p from 0 to |A| step 1 do
    if labels_p = index then
        APPEND(active, p)
    else
        APPEND(inactive, p)
thresh ← THRESHOLD()
if |centroid_0 – centroid_1| < thresh then
    <centroid_0, centroid_1, labels, index> ← KMEANS(inactive)
    for q from 0 to |inactive| step 1 do
        if labels_q = index then
            APPEND(active, inactive_q)
return <active>
```

KMEANS(A)

C ← empty-list
C[0] ← max(A)
C[1] ← min(A)
K ← kmeans(2, centroids=C) [53]
clusters ← K.clusters(A)
labels ← K.labels(A)
centroids ← clusters.centroid
index ← 0 if centroids[0] < centroids[1]
    index ← 1
return <centroids[0],centroids[1],labels, index>

THRESHOLD()

thresholds ← empty-list
thresh ← 0
foreach motion d in FULL-DATASET do
    SD ← STANDARD-DEVIATION(d)
    <$centroid_0$, $centroid_1$, labels, index> ← KMEANS(SD)
    difference ← | $centroid_0$ – $centroid_1$ |
    APPEND(thresholds, difference)
average ← SUM(thresholds)/|FULL-DATASET|
return <average>

Step 2: Apply the normalization step. This step is subdivided into four steps: resampling, rotation, scaling, restore-orientation, and scaling. These steps were adapted from the $3 stroke gesture recognition algorithm [69], protractor3D [70], and protractor [76].

NORMALIZATION(C, T)

C' ← C
p ← 32
foreach joint j^i for i ≥ 0 in C do
    norm_j^i ← RESAMPLE(j^i, p) [69]
    norm_j^i ← ROTATE-TO-ORIGIN(norm_j^i) [69, 76]
    norm_j^i ← SCALE(norm_j^i) [69]
    norm_j^i ← TRANSLATE-TO-ORIGIN(norm_j^i) [69]
    APPEND(C', j^i)
T' ← T
foreach motion t in T do
    t' ← t
    foreach joint j^i in t do
        norm_j^i ← RESAMPLE(j^i, p)
        norm_j^i ← ROTATE-TO-ORIGIN(norm_j^i)

```
            norm_j^i ← SCALE(norm_j^i)
            norm_j^i ← TRANSLATE-TO-ORIGIN(norm_j^i)
            APPEND(t', norm_j^i)
        APPEND(T', norm_j^i)
return < C', T'>
```

RECOGNIZE(C', T')

```
b ← +inf
foreach motion t' in T' do
        d ← 0
        foreach joint j^i_C', j^i_t' in C', T' respectively do
            r_t ← OPTIMAL-ALIGNMENT(j^i_C', j^i_t') [69]
            r_c ← OPTIMAL-ALIGNMENT(j^i_t', j^i_C') [69]
            flip_j^i ← FLIP(j^i_C')
            flip_r_t ← OPTIMAL-ALIGNMENT(flip_j^i, j^i_t')
            flip_r_c ← OPTIMAL-ALIGNMENT(j^i_t', flip_j^i)
            e_1 ← PATH-DISTANCE(j^i_C', r_t )
            e_2 ← PATH-DISTANCE(r_c, j^i_t')
            e_3 ← PATH-DISTANCE(flip_j^i, flip_r_t )
            e_4 ← PATH-DISTANCE(flip_r_c, j^i_t')
            d+= min(e_1, e_2, e_3, e_4)
        if d < b then
            b ← d
            t'' ← t'
return < t''>
```

FLIP(A)

```
for i from 0 to |A| step 1 do
        A[i]_x ← - A[i]_x
return <A>
```

PATH-DISTANCE(A, B)

```
d ← 0
for i from 0 to |A| step 1 do
        d += EUCLIDEAN-DISTANCE(A_i, B_i)
return <d>
```

APPENDIX F
FILTERJOINT METHOD RESULTS

Table F-1. List of Actively Moving Joints Selected by our filterJoint Method. Adults are more consistent than children especially for motions that require less joints to perform the movement (e.g., raise your hand and raise your arm to one side)

ID	Joints Selected
	Bend your knee
	Adult
565	FootLeft, KneeLeft, HandLeft, WristLeft, ElbowLeft, HandRight, ShoulderLeft, AnkleLeft, Head, WristRight, ShoulderRight, ShoulderCenter, HipRight, HipCenter, HipLeft, Spine
577	KneeRight, FootRight, AnkleRight, HandLeft, WristLeft, ElbowLeft, KneeLeft, Head, HandRight, ShoulderLeft
604	KneeRight, FootRight, AnkleRight, HandLeft, WristLeft, ElbowLeft, ShoulderRight, ElbowRight, WristRight, Head, HandRight, ShoulderCenter, ShoulderLeft, Spine, HipRight, HipCenter, HipLeft
734	KneeRight, AnkleRight, FootRight, HandRight, WristRight, ElbowRight, HipRight, ShoulderRight
876	KneeRight, FootRight, AnkleRight, HandRight, Head, ShoulderLeft, HipRight, ShoulderRight, HipCenter, Spine, HipLeft, ShoulderCenter, ElbowRight, WristRight
888	KneeRight, FootRight, AnkleRight, HandRight, FootLeft, HipRight, HipLeft, HipCenter, WristRight, Spine, ShoulderRight
921	FootRight, KneeRight, AnkleRight, HandRight, WristRight
934	FootRight, AnkleRight, KneeRight, WristLeft, HandLeft, ElbowLeft, HipRight, Head, ShoulderLeft, HipCenter, Spine, ShoulderCenter, HipLeft, ShoulderRight, ElbowRight, HandRight
970	KneeRight, AnkleRight, FootRight
976	FootRight, AnkleRight, KneeRight, HandLeft, Head, WristLeft, ShoulderLeft, ShoulderRight, ShoulderCenter, ElbowLeft, ElbowRight, Spine, HipCenter, HipLeft, HipRight, WristRight, HandRight
	Child
103	KneeRight, AnkleRight, FootRight, HandRight, WristRight, ElbowRight, ShoulderCenter, HandLeft, WristLeft, HipLeft, ShoulderRight, ShoulderLeft, Head, HipCenter
106	HandRight, WristRight, ElbowRight, FootRight, AnkleRight
169	FootRight, AnkleRight, KneeRight, HipRight, Head, HipLeft, HipCenter, WristRight, HandLeft, WristLeft, Spine, ElbowRight, HandRight, ShoulderLeft, ShoulderCenter
290	FootRight, KneeRight, AnkleRight, Head, ShoulderRight, ShoulderCenter, HandRight, ShoulderLeft, HipLeft, HipRight, Spine, HipCenter, ElbowRight, WristRight, HandLeft, KneeLeft, WristLeft, ElbowLeft
337	HandLeft, WristLeft, KneeRight, AnkleRight, FootRight, ElbowLeft, HandRight, WristRight, ShoulderLeft, ElbowRight, Head, ShoulderRight, ShoulderCenter, HipCenter, HipLeft, HipRight, Spine

Table F-1. Continued

ID	Joints Selected
342	HandRight, HandLeft, WristLeft, WristRight
474	AnkleRight, FootRight, KneeRight, WristRight, HandRight, FootLeft, Head, HandLeft, WristLeft
595	KneeRight, AnkleRight, FootRight, HandRight, HandLeft, WristLeft, WristRight, ElbowLeft, ElbowRight, ShoulderRight, Head
644	FootRight, KneeRight, AnkleRight, HandLeft, WristLeft, HandRight, WristRight, ElbowLeft, ElbowRight, HipLeft, ShoulderLeft, HipCenter, HipRight, Spine, Head, ShoulderCenter, ShoulderRight, KneeLeft
723	KneeRight, AnkleRight, FootRight, HandRight, WristRight, HandLeft, WristLeft, ElbowRight
	Bow
	Adult
565	Head, ShoulderCenter, ShoulderLeft, HandLeft, ShoulderRight, HandRight, HipRight, WristRight, HipLeft, Spine, HipCenter, WristLeft, ElbowRight, ElbowLeft, KneeLeft, KneeRight
577	Head, HandRight, ShoulderCenter, WristRight, HandLeft, ShoulderRight, WristLeft, ShoulderLeft, HipRight, HipLeft, HipCenter, KneeLeft, KneeRight, Spine, ElbowRight, ElbowLeft
604	Head, ShoulderCenter, ShoulderLeft, ShoulderRight, HipLeft, HipRight, Spine, HipCenter, HandLeft, WristLeft, HandRight, KneeRight, WristRight, KneeLeft, ElbowLeft
734	Head, ShoulderCenter, ShoulderLeft, ShoulderRight, HipLeft, HipRight, Spine, HipCenter, ElbowLeft, WristLeft, HandLeft, ElbowRight, HandRight, WristRight
876	Head, ShoulderCenter, HandRight, ShoulderRight, ShoulderLeft, WristRight, HandLeft, HipLeft, HipRight, Spine, HipCenter, WristLeft, ElbowRight, KneeRight, KneeLeft, ElbowLeft
888	Head, ShoulderCenter, ShoulderRight, ShoulderLeft, HipLeft, HipRight, HandRight, WristRight, HipCenter, KneeLeft, WristLeft, HandLeft, Spine, KneeRight
921	Head, ShoulderCenter, ShoulderRight, ShoulderLeft, HipRight, ElbowLeft, Spine, HipLeft, HipCenter, ElbowRight, HandLeft, WristLeft, KneeRight, WristRight
934	Head, ShoulderCenter, ShoulderRight, ShoulderLeft, HandRight, WristRight, HandLeft, HipRight, HipLeft, ElbowRight, HipCenter, Spine, WristLeft, KneeLeft, KneeRight
970	Head, ShoulderCenter, ShoulderRight, ShoulderLeft, HandLeft, WristLeft, HipLeft, HipRight, HipCenter, Spine, ElbowLeft, HandRight, WristRight, ElbowRight, KneeRight, KneeLeft

Table F-1. Continued

ID	Joints Selected
976	Head, ShoulderCenter, ShoulderRight, ShoulderLeft, HandLeft, HandRight, HipRight, Spine, HipLeft, HipCenter, WristLeft, WristRight, KneeLeft, KneeRight, ElbowRight, ElbowLeft
	Child
103	Head, ShoulderCenter, HandLeft, ShoulderLeft, AnkleLeft, WristLeft, ShoulderRight, FootLeft, KneeLeft, HandRight, ElbowLeft, WristRight, HipLeft, Spine, HipRight, HipCenter
106	HandRight, Head, WristRight, ShoulderCenter, ElbowRight, HandLeft, WristLeft, HipLeft, ElbowLeft, HipRight, ShoulderRight, HipCenter, Spine, ShoulderLeft, KneeLeft, KneeRight
169	Head, ShoulderCenter, HandLeft, WristLeft, HipRight, HipLeft, HandRight, ShoulderLeft, HipCenter, WristRight, Spine, ElbowLeft, ShoulderRight, ElbowRight, KneeRight, KneeLeft
290	HandRight, WristRight, HandLeft, WristLeft, ElbowRight, Head, ShoulderCenter
337	HandLeft, WristLeft, Head, ShoulderCenter, ElbowLeft, ShoulderLeft, HandRight, WristRight, ShoulderRight, ElbowRight, Spine, HipCenter, HipLeft, HipRight
342	WristLeft, HandLeft, Head, ShoulderCenter, ShoulderRight, WristRight, HandRight, ElbowLeft, ElbowRight, HipRight, HipLeft, ShoulderLeft, HipCenter, Spine, KneeRight
474	Head, ShoulderCenter, HipRight, ShoulderRight, HandLeft, HipLeft, HandRight, Spine, HipCenter, WristRight, ShoulderLeft, WristLeft, ElbowRight, ElbowLeft, KneeLeft, KneeRight
595	Head, ShoulderCenter, HandLeft, WristLeft, ShoulderRight, HipRight, HipLeft, Spine, HipCenter, ShoulderLeft, KneeLeft, HandRight, ElbowLeft, WristRight, ElbowRight
644	HandRight, Head, ShoulderCenter, WristRight, ShoulderLeft, HandLeft, ShoulderRight, WristLeft, ElbowLeft, HipLeft, HipRight, Spine, HipCenter, ElbowRight
723	Head, ShoulderCenter, ShoulderRight, ShoulderLeft, Spine, HipCenter, HipLeft, HipRight, HandRight, WristRight, HandLeft, WristLeft, KneeLeft, KneeRight, ElbowRight, ElbowLeft
	Do a forward lunge
	Adult
565	FootRight, AnkleRight, KneeRight, HandRight, Head, ShoulderCenter, ShoulderRight, WristRight, ElbowRight, ShoulderLeft, Spine, ElbowLeft, HandLeft, WristLeft, HipCenter, HipRight, HipLeft
577	FootRight, AnkleRight, KneeRight, HandLeft, HandRight, WristLeft, WristRight, ShoulderRight, ElbowLeft, ShoulderCenter, ElbowRight, Head, ShoulderLeft, Spine, HipLeft, HipCenter, HipRight

Table F-1. Continued

ID	Joints Selected
604	HandLeft, WristLeft, KneeRight, FootRight, AnkleRight, ShoulderLeft, Head, ShoulderCenter, ElbowLeft, ShoulderRight, Spine, HipLeft, HipCenter, FootLeft, HandRight, HipRight
734	KneeRight, Head, FootRight, ShoulderCenter, AnkleRight, ShoulderLeft, ShoulderRight, Spine, ElbowRight, ElbowLeft, HandRight, WristRight, HipCenter
876	FootRight, AnkleRight, KneeRight, HandRight, Head, WristRight, ShoulderCenter, HandLeft, ShoulderLeft, ShoulderRight, ElbowLeft, WristLeft, ElbowRight, Spine
888	KneeRight, FootRight, AnkleRight, HandRight, WristRight, ElbowRight, ShoulderRight, Head, ShoulderCenter, HipRight, Spine, HipCenter, HipLeft, ShoulderLeft, KneeLeft, ElbowLeft
921	KneeRight, ElbowLeft, WristLeft, ShoulderLeft, HandLeft, Head, ShoulderCenter, ShoulderRight, Spine, WristRight, AnkleRight, HipLeft, HandRight, HipCenter, ElbowRight, FootRight, HipRight
934	KneeLeft, FootLeft, AnkleLeft, Head, ShoulderLeft, ShoulderCenter, Spine, ShoulderRight, ElbowLeft, ElbowRight, HipCenter, WristRight, HandRight, HipLeft, HipRight, WristLeft, HandLeft
970	KneeRight, AnkleRight, FootRight, HandRight, WristRight, HandLeft, ElbowRight, WristLeft, ShoulderCenter, ShoulderRight, Head, ShoulderLeft, ElbowLeft, Spine, HipCenter, HipLeft, HipRight
976	KneeRight, Head, ShoulderCenter, FootRight, AnkleRight, ShoulderLeft, ElbowLeft, ShoulderRight, Spine, WristLeft, HipLeft, HipCenter, HipRight, HandLeft, WristRight, ElbowRight, HandRight
	Child
103	Head, ShoulderCenter, ShoulderRight, ShoulderLeft, Spine, ElbowLeft, KneeLeft, ElbowRight, HipCenter, FootLeft, AnkleLeft, HipLeft, HipRight, WristLeft, HandLeft, WristRight, HandRight
106	FootRight, AnkleRight, KneeRight, HandRight, WristRight, ElbowRight, HipRight, HipLeft, HipCenter, ShoulderRight, Spine, ShoulderCenter, ElbowLeft, WristLeft, ShoulderLeft, HandLeft, Head, KneeLeft
169	KneeRight, FootRight, AnkleRight, ShoulderLeft, Head, ShoulderCenter, ElbowLeft, Spine, HipLeft, WristLeft, ShoulderRight, HipCenter, HandLeft
290	AnkleLeft, FootLeft AnkleLeft, FootLeft, HandLeft, Head, AnkleRight, WristLeft, KneeLeft, FootRight, ShoulderLeft, KneeRight, ElbowLeft
337	HandRight, HandLeft, WristRight, WristLeft, ElbowRight, ElbowLeft
342	HandLeft, WristLeft, AnkleLeft, KneeLeft, ElbowLeft, FootLeft, ElbowRight, HandRight, ShoulderLeft, WristRight, ShoulderCenter, ShoulderRight, Head, Spine, HipLeft, HipCenter, HipRight, FootRight
474	ShoulderCenter, Head, ShoulderLeft, KneeRight, ShoulderRight, AnkleLeft, ElbowLeft, Spine, HandLeft, AnkleRight, WristRight, HandRight, WristLeft, HipCenter, FootRight, HipLeft, HipRight, ElbowRight, FootLeft

Table F-1. Continued

ID	Joints Selected
595	HandLeft, WristLeft, ElbowRight, ShoulderRight, AnkleLeft, Head, ShoulderCenter, HandRight, ShoulderLeft, WristRight, ElbowLeft, Spine, FootLeft, KneeRight, HipRight, HipCenter, HipLeft
644	FootRight, AnkleRight, KneeRight, Head, ShoulderCenter, ShoulderRight, Spine, ShoulderLeft, ElbowRight, HipCenter, HipRight, HipLeft, ElbowLeft, HandLeft, WristRight, WristLeft, HandRight
723	Head, ShoulderRight, ShoulderCenter, KneeRight, ShoulderLeft, ElbowRight, Spine, WristRight, HandRight, HipCenter, HipLeft, HipRight, ElbowLeft, AnkleRight, WristLeft, FootRight, HandLeft, AnkleLeft, KneeLeft
	Jump
	Adult
565	HandRight, WristRight, ElbowRight, ShoulderRight, ElbowLeft, ShoulderLeft, Head, ShoulderCenter, HandLeft, KneeLeft, Spine, HipRight, WristLeft, HipCenter, KneeRight, HipLeft, AnkleRight, AnkleLeft
577	KneeRight, HandLeft, KneeLeft, WristLeft, ShoulderLeft, ShoulderCenter, Head, ElbowRight, ShoulderRight, ElbowLeft, AnkleRight, Spine, HandRight, WristRight, HipCenter, HipLeft, HipRight, FootRight
604	HandLeft, WristLeft, Head, ShoulderCenter, HandRight, ElbowLeft, ShoulderLeft, WristRight, ShoulderRight, ElbowRight
734	Head, ShoulderCenter, ShoulderRight, ShoulderLeft, ElbowRight, HandRight, ElbowLeft, WristRight, HandLeft, WristLeft, Spine, HipLeft, HipCenter, HipRight, KneeLeft, KneeRight, AnkleRight, AnkleLeft
876	HandRight, HandLeft, WristRight, KneeLeft, WristLeft, KneeRight, ShoulderLeft, ShoulderCenter, Head, ElbowRight, ShoulderRight, ElbowLeft
888	HandLeft, WristLeft, HandRight, KneeRight, ElbowLeft, WristRight, KneeLeft, ShoulderLeft, ShoulderCenter, Head, ElbowRight
921	HandLeft, WristLeft, ElbowLeft, HandRight, WristRight, KneeRight, ShoulderLeft, ShoulderCenter, KneeLeft, ShoulderRight, Head, ElbowRight, Spine, HipRight, HipCenter, HipLeft
934	WristLeft, HandLeft, Head, ElbowLeft, ShoulderCenter, HandRight, ShoulderLeft, ShoulderRight, ElbowRight, WristRight, Spine, HipLeft, HipCenter, HipRight, KneeLeft, KneeRight
970	HandRight, HandLeft, WristRight, WristLeft, ElbowLeft, ElbowRight, KneeLeft, Head, ShoulderCenter, ShoulderRight, Spine, ShoulderLeft
976	HandLeft, HandRight, WristLeft, WristRight
	Child
103	HandRight, WristRight, HandLeft, WristLeft, ElbowRight, ElbowLeft, HipLeft, ShoulderRight, HipCenter, ShoulderLeft, HipRight, Spine, Head, ShoulderCenter
106	HandRight, WristRight, HandLeft, WristLeft, ElbowRight, ElbowLeft, Head, FootRight, AnkleRight, ShoulderCenter, Spine, ShoulderLeft, ShoulderRight, HipCenter, HipLeft, HipRight

Table F-1. Continued

ID	Joints Selected
169	HandRight, WristRight, KneeLeft, ElbowRight, HandLeft, KneeRight, WristLeft, FootLeft, FootRight, ElbowLeft, AnkleLeft, AnkleRight
290	HandRight, WristRight, HandLeft, FootRight, WristLeft, AnkleRight, ElbowRight, ElbowLeft, KneeRight, Spine, ShoulderLeft, HipRight, ShoulderRight, ShoulderCenter, HipCenter, HipLeft, Head
337	HandLeft, WristLeft, FootRight, AnkleRight, KneeRight, HandRight, ShoulderRight, Head
342	HandRight, HandLeft, WristRight, WristLeft, ElbowRight, ElbowLeft, Head
474	HandLeft, WristLeft, HandRight, WristRight, ElbowRight, ElbowLeft, ShoulderCenter, ShoulderRight, Head, ShoulderLeft, Spine, HipCenter, HipLeft, HipRight, KneeRight, KneeLeft
595	HandRight, HandLeft, WristRight, WristLeft, ElbowLeft, ElbowRight, ShoulderLeft, FootRight
644	HandRight, WristRight, HandLeft, WristLeft, ElbowRight, ElbowLeft, Head, ShoulderRight, ShoulderLeft, HipLeft, Spine, KneeLeft, HipCenter, HipRight, ShoulderCenter, KneeRight
723	HandRight, WristRight, HandLeft, ElbowRight, WristLeft, FootRight, AnkleRight, ElbowLeft, ShoulderRight, HipRight, Spine, Head, HipCenter, HipLeft, ShoulderCenter, KneeRight, ShoulderLeft, KneeLeft
	Kick a ball as hard as you can
	Adult
565	HandRight, WristRight, AnkleRight, FootRight, KneeRight
577	FootRight, AnkleRight, HandLeft, WristLeft, KneeRight
604	FootRight, HandLeft, AnkleRight, WristLeft, ElbowLeft, KneeRight, HandRight, WristRight, ShoulderRight, ShoulderLeft, HipRight, ElbowRight, HipCenter, HipLeft, Spine, ShoulderCenter, Head
734	HandLeft, WristLeft, FootRight, AnkleRight
876	FootRight, AnkleRight, KneeRight, HandLeft, WristLeft, FootLeft, AnkleLeft, HandRight, KneeLeft, WristRight
888	FootRight, AnkleRight, KneeRight
921	FootRight, AnkleRight, KneeRight
934	FootRight, HandLeft, AnkleRight, WristLeft, KneeRight, AnkleLeft, FootLeft, ElbowLeft, KneeLeft
970	FootRight, AnkleRight, HandRight, KneeRight, WristRight, HandLeft, ElbowRight, AnkleLeft, FootLeft, WristLeft, KneeLeft, HipRight, ElbowLeft, HipLeft, ShoulderRight, HipCenter
976	FootRight, AnkleRight, HandLeft, KneeRight, WristLeft, HandRight, WristRight, ElbowRight
	Child
103	HandRight, HandLeft, FootRight, WristRight, WristLeft, AnkleRight, KneeRight, ElbowRight, ElbowLeft, Head, ShoulderLeft, ShoulderCenter, ShoulderRight, HipRight, HipLeft, HipCenter, Spine, KneeLeft
106	FootRight, AnkleRight, HandRight, KneeRight, WristRight, HandLeft, ElbowRight, WristLeft, HipRight

Table F-1. Continued

ID	Joints Selected
169	FootRight, AnkleRight, HandRight, WristRight, KneeRight, ElbowRight
290	FootRight, AnkleRight, KneeRight, HandRight, WristRight, ElbowRight, ShoulderRight, HandLeft, WristLeft, HipRight, Spine, Head, ShoulderCenter, HipCenter
337	FootRight, AnkleRight, HandLeft, HandRight, WristLeft, WristRight, ElbowLeft, ShoulderLeft, ElbowRight, ShoulderCenter, KneeRight, ShoulderRight, Head, Spine, HipCenter, HipRight
342	HandRight, WristRight, HandLeft, WristLeft, FootRight, AnkleRight, ElbowRight, KneeRight, ElbowLeft
474	FootRight, AnkleRight, HandLeft, WristLeft, HandRight, WristRight, ElbowLeft, ElbowRight, KneeRight, ShoulderLeft, ShoulderCenter, ShoulderRight, Head, Spine
595	AnkleRight, FootRight, KneeRight, Head, HandRight, WristRight, ShoulderRight, ElbowRight, ShoulderCenter, HandLeft
644	FootRight, AnkleRight, HandLeft, WristLeft, KneeRight, ElbowLeft, HandRight, Head, WristRight, ShoulderRight, ShoulderCenter, ShoulderLeft, ElbowRight
723	FootRight, AnkleRight, HandRight, WristRight, KneeRight, ElbowRight, HipRight, ShoulderRight
	Lift your leg to one side
	Adult
565	FootRight, AnkleRight, KneeRight, HandRight, WristRight, HandLeft, WristLeft, ShoulderCenter, ElbowLeft, ElbowRight, Head, ShoulderLeft, Spine, ShoulderRight, HipCenter, HipRight, HipLeft, KneeLeft
577	FootRight, AnkleRight, KneeRight
604	FootRight, AnkleRight
734	FootRight, AnkleRight, KneeRight
876	FootRight, AnkleRight, KneeRight
888	FootRight, AnkleRight, KneeRight, HandLeft, WristLeft, HandRight, ShoulderCenter, ShoulderRight, ShoulderLeft, ElbowLeft, Head, WristRight
921	FootRight, AnkleRight, KneeRight, ShoulderCenter, Head, ShoulderLeft, ShoulderRight, ElbowLeft, WristLeft, Spine, HandLeft, ElbowRight, HipCenter, HipLeft, HipRight
934	FootRight, AnkleRight
970	FootRight, AnkleRight, KneeRight
976	FootRight, AnkleRight, HandLeft
	Child
103	FootRight, AnkleRight, KneeRight, HandLeft, WristLeft, ElbowLeft, Head, ShoulderCenter
106	FootRight, AnkleRight
169	FootRight, AnkleRight, HandLeft, WristLeft, KneeRight, HandRight, WristRight

Table F-1. Continued

ID	Joints Selected
290	FootRight, AnkleRight, HandLeft, HandRight, WristRight, WristLeft, KneeRight, Head, ElbowLeft, ElbowRight, ShoulderCenter, ShoulderRight, ShoulderLeft, Spine, HipRight, HipCenter
337	HandLeft, WristLeft, ElbowLeft, FootRight, HandRight, AnkleRight, WristRight, ElbowRight, KneeRight, ShoulderLeft, ShoulderRight, ShoulderCenter, Head
342	FootRight, AnkleRight, HandRight, WristRight, KneeRight, ElbowRight
474	FootRight, AnkleRight, HandLeft, WristLeft, KneeRight, ElbowLeft
595	FootRight, AnkleRight, KneeRight
644	FootRight, AnkleRight, KneeRight
723	FootRight, AnkleRight, KneeRight
	Point at the camera
	Adult
565	HandRight, WristRight
577	HandRight, WristRight
604	HandRight, WristRight, ElbowRight
734	HandRight, WristRight
876	HandRight, WristRight
888	HandRight, WristRight
921	HandRight, WristRight
934	HandRight, WristRight
970	HandRight, WristRight, ElbowRight
976	HandRight, WristRight, ElbowRight
	Child
103	HandRight, WristRight, HandLeft, WristLeft, ElbowRight, ElbowLeft
106	HandRight, WristRight, ElbowRight, ShoulderRight, ElbowLeft
169	HandRight, WristRight, ElbowRight, ShoulderRight, KneeLeft, HandLeft, WristLeft, ElbowLeft, ShoulderLeft
290	HandRight, WristRight, ElbowRight, HandLeft, WristLeft, ShoulderRight
337	HandLeft, WristLeft, HandRight, WristRight, ElbowLeft, ElbowRight, ShoulderRight, ShoulderLeft, Head, ShoulderCenter
342	HandRight, WristRight, HandLeft, WristLeft, ElbowLeft, ElbowRight, ShoulderLeft, ShoulderRight
474	HandRight, WristRight
595	HandRight, WristRight
644	HandRight, WristRight, ElbowRight
723	HandRight, WristRight, ElbowRight
	Punch
	Adult
565	HandRight, WristRight
577	HandRight, WristRight
604	HandLeft, HandRight, WristRight, WristLeft, FootRight, AnkleRight, KneeRight, ElbowRight, HipRight, ShoulderRight, HipCenter, Head, Spine, ShoulderCenter

Table F-1. Continued

ID	Joints Selected
734	HandRight, WristRight, HandLeft, ElbowRight, WristLeft
876	HandRight, WristRight, HandLeft, ElbowRight
888	HandRight, WristRight
921	HandLeft, WristLeft, ElbowLeft
934	HandRight, WristRight, ElbowRight
970	HandRight, WristRight, HandLeft, ElbowRight, WristLeft, AnkleLeft, FootLeft, ShoulderRight, HipRight, KneeLeft, ShoulderCenter, Head, HipCenter, ElbowLeft, Spine, ShoulderLeft, HipLeft
976	HandRight, WristRight, HandLeft, WristLeft, ElbowRight, ElbowLeft, ShoulderRight, FootLeft, Head, AnkleLeft, ShoulderCenter, Spine, ShoulderLeft, KneeLeft, HipRight, HipCenter
	Child
103	HandRight, HandLeft, WristRight, WristLeft, ElbowRight
106	HandLeft, WristLeft, HandRight, WristRight, ElbowLeft, ElbowRight, ShoulderLeft, ShoulderRight
169	HandRight, WristRight, ElbowRight
290	HandRight, WristRight, ElbowRight, HandLeft, WristLeft
337	HandRight, WristRight, FootRight, AnkleRight, ElbowRight, KneeRight, WristLeft, HandLeft
342	HandRight, WristRight, ElbowRight
474	HandRight, WristRight
595	HandRight, WristRight, FootLeft, AnkleLeft, HandLeft, ElbowRight, WristLeft, ShoulderRight, Head, ElbowLeft, KneeLeft, ShoulderCenter, ShoulderLeft, Spine, HipRight, HipCenter, HipLeft
644	HandRight, WristRight, HandLeft, WristLeft, ElbowRight, ShoulderRight, KneeRight, ElbowLeft, Head
723	HandRight, WristRight, HandLeft, WristLeft, ElbowRight, ElbowLeft, AnkleRight, ShoulderRight, FootRight, HipRight, ShoulderCenter, Spine, KneeRight, Head, HipCenter
	Put your hands on your hips and lean to the side
	Adult
565	HandRight, WristRight, Head, ElbowRight, ShoulderCenter, ShoulderLeft, ShoulderRight
577	Head, HandRight, HandLeft, WristLeft, ElbowLeft, WristRight, ShoulderCenter, ElbowRight, ShoulderLeft, ShoulderRight, HipRight, Spine, KneeRight, HipCenter, KneeLeft
604	Head, HandRight, ShoulderLeft, ShoulderCenter, WristRight, ShoulderRight, ElbowRight, HandLeft, ElbowLeft, WristLeft, Spine, HipCenter, HipLeft, HipRight
734	Head, ShoulderLeft, ElbowLeft, ShoulderCenter, HandLeft, WristLeft, HandRight, ElbowRight, ShoulderRight, WristRight, HipRight, HipCenter, HipLeft, Spine, KneeLeft, KneeRight

Table F-1. Continued

ID	Joints Selected
876	Head, HandRight, ShoulderCenter, HandLeft, ShoulderLeft, ElbowLeft, WristLeft, WristRight, ShoulderRight, ElbowRight, HipRight, Spine, HipCenter, FootRight, HipLeft
888	Head, HandRight, ElbowLeft, HandLeft, WristLeft, WristRight, ShoulderLeft, ShoulderCenter, ElbowRight, ShoulderRight, HipRight
921	Head, ShoulderCenter, ShoulderLeft, ShoulderRight, ElbowLeft, HandLeft, ElbowRight, WristLeft, WristRight, HandRight, HipRight, Spine, HipCenter, KneeLeft
934	Head, ShoulderRight, HandRight, ShoulderCenter, ShoulderLeft, ElbowRight, ElbowLeft, HandLeft, WristRight, WristLeft, HipRight, Spine, HipCenter, HipLeft, KneeLeft, KneeRight
970	HandRight, Head, WristRight, ShoulderCenter, ShoulderRight, ElbowRight, HandLeft, ShoulderLeft, ElbowLeft, WristLeft, Spine, HipRight, HipCenter, HipLeft
976	HandLeft, WristLeft, ElbowLeft, HandRight, WristRight, HipRight, KneeLeft, AnkleLeft, HipLeft, HipCenter, ElbowRight, FootLeft, Spine, ShoulderRight, KneeRight, Head
	Child
103	Head, WristLeft, HandLeft, ElbowLeft, ShoulderCenter, ShoulderRight, ElbowRight, ShoulderLeft, WristRight, HandRight, Spine, FootLeft, HipCenter, HipLeft
106	HandRight, WristRight, HandLeft, WristLeft, ElbowLeft, Head, ElbowRight, ShoulderRight, ShoulderCenter, ShoulderLeft
169	Head, ElbowLeft, HandLeft, HandRight, WristLeft, WristRight, ShoulderCenter, ShoulderLeft, ShoulderRight, ElbowRight, HipRight, HipCenter, Spine, KneeRight, HipLeft, KneeLeft
290	HandRight, WristRight, Head, ElbowRight, HandLeft, ElbowLeft, ShoulderRight, WristLeft, ShoulderLeft, ShoulderCenter, HipRight, KneeLeft, HipCenter, Spine, KneeRight, HipLeft
337	HandLeft, WristLeft, ElbowRight, WristRight, HandRight, ElbowLeft, ShoulderRight, Head, ShoulderCenter, ShoulderLeft, Spine, HipCenter, HipRight, HipLeft, FootLeft, AnkleLeft
342	Head, HandLeft, WristLeft, WristRight, HandRight, ElbowLeft, ShoulderRight, ShoulderCenter, ElbowRight, ShoulderLeft, KneeRight, HipRight, Spine, HipCenter, HipLeft
474	HandRight, Head, WristRight, HandLeft, ElbowLeft, WristLeft, ShoulderCenter, ShoulderLeft, ShoulderRight, ElbowRight
595	Head, ShoulderCenter, HipLeft, ShoulderLeft, WristLeft, HandLeft, WristRight, HipRight, HipCenter, HandRight, ElbowLeft, ShoulderRight, ElbowRight, Spine, KneeLeft, KneeRight, FootLeft, AnkleLeft

Table F-1. Continued

ID	Joints Selected
644	Head, HandLeft, WristLeft, ElbowLeft, ShoulderCenter, ShoulderRight, ElbowRight, ShoulderLeft, HandRight, WristRight, Spine, HipRight, HipCenter, HipLeft
723	Head, ShoulderLeft, ShoulderCenter, HandRight, HandLeft, WristRight, WristLeft, ElbowLeft, ShoulderRight, ElbowRight, KneeRight, HipRight, Spine, HipCenter, HipLeft
	Raise your arm to one side
	Adult
565	HandRight, WristRight
577	HandRight, WristRight
604	HandRight, WristRight
734	HandRight, WristRight
876	HandRight, WristRight
888	HandRight, WristRight
921	HandRight, WristRight
934	HandRight, WristRight
970	HandRight, WristRight
976	HandRight, WristRight
	Child
103	HandRight, WristRight, ElbowRight, HandLeft, ShoulderRight, FootLeft, WristLeft, KneeRight, Head
106	HandRight, WristRight, ElbowRight
169	HandRight, WristRight, ElbowRight
290	HandRight, WristRight, HandLeft, WristLeft, ElbowRight
337	HandRight, WristRight, ElbowRight
342	HandRight, WristRight
474	HandRight, WristRight, ElbowRight
595	HandRight, WristRight, ElbowRight
644	HandRight, WristRight
723	HandRight, WristRight, ElbowRight
	Raise your hand
	Adult
565	HandRight, WristRight, ElbowRight
577	HandRight, WristRight
604	HandRight, WristRight, ElbowRight
734	HandRight, WristRight
876	HandRight, WristRight
888	HandRight, WristRight
921	HandRight, WristRight
934	HandRight, WristRight
970	HandRight, WristRight, ElbowRight
976	HandRight, WristRight

Table F-1. Continued

ID	Joints Selected
	Child
103	HandRight, WristRight, ElbowRight
106	HandRight, WristRight, ElbowRight
169	HandRight, WristRight, ElbowRight
290	HandRight, WristRight, ElbowRight
337	HandRight, WristRight, ElbowRight
342	HandRight, WristRight, ElbowRight
474	HandRight, HandLeft, WristRight, WristLeft, ElbowRight, ElbowLeft
595	HandRight, WristRight, ElbowRight
644	HandRight, WristRight, ElbowRight
723	HandRight, WristRight, ElbowRight
	Swipe across an imaginary screen in front of you
	Adult
565	HandRight, WristRight, ElbowRight
577	HandRight, WristRight
604	HandRight, WristRight, ElbowRight
734	HandRight, WristRight
876	HandRight, WristRight
888	HandRight, WristRight
921	HandRight, WristRight
934	HandRight, WristRight
970	HandRight, WristRight, ElbowRight, HandLeft, WristLeft
976	HandRight, WristRight
	Child
103	HandRight, WristRight, ElbowRight
106	HandRight, WristRight, ElbowRight, ShoulderRight, ElbowLeft, Head, WristLeft, HandLeft, ShoulderCenter, ShoulderLeft, FootLeft, AnkleLeft, KneeLeft
169	HandRight, WristRight, ElbowRight
290	HandRight, WristRight, ElbowRight, HandLeft, WristLeft, ElbowLeft, ShoulderLeft, Head, ShoulderRight
337	HandRight, WristRight, ElbowRight, HandLeft, WristLeft
342	HandRight, WristRight
474	HandRight, WristRight
595	HandRight, WristRight, ElbowRight, Head, ShoulderRight
644	HandRight, WristRight
723	HandRight, WristRight, ElbowRight
	Throw a ball as far as you can
	Adult
565	HandRight, WristRight, ElbowRight
577	HandRight, WristRight, ElbowRight, HandLeft, WristLeft, ShoulderRight, Head, ShoulderCenter, ShoulderLeft, ElbowLeft, Spine, HipCenter, HipRight

Table F-1. Continued

ID	Joints Selected
604	HandRight, WristRight, HandLeft, ElbowRight, WristLeft, ShoulderRight, Head, ElbowLeft, ShoulderCenter, ShoulderLeft, Spine, FootLeft, AnkleLeft, HipRight, HipCenter
734	HandRight, WristRight, ElbowRight, HandLeft
876	HandRight, WristRight, ElbowRight
888	HandRight, WristRight, ElbowRight, HandLeft, WristLeft
921	HandRight, WristRight
934	HandRight, WristRight, ElbowRight
970	HandLeft, ElbowRight, FootLeft, AnkleLeft, HandRight, WristRight, WristLeft, ShoulderRight, KneeLeft, ShoulderCenter, Head, ElbowLeft, ShoulderLeft, Spine, HipRight, HipCenter, HipLeft
976	HandRight, WristRight, HandLeft, WristLeft, ElbowRight
	Child
103	HandRight, WristRight, ElbowRight
106	HandRight, WristRight, ElbowRight, HandLeft, ShoulderRight, WristLeft, FootRight, KneeRight, AnkleRight
169	HandRight, WristRight, ElbowRight, HandLeft, WristLeft
290	HandRight, HandLeft, WristRight, WristLeft
337	HandRight, WristRight, HandLeft, WristLeft, ElbowRight, ElbowLeft, ShoulderLeft, ShoulderRight, Head, HipLeft, ShoulderCenter, HipCenter, HipRight, AnkleRight, FootRight, Spine, KneeRight
342	Head, HandRight, ElbowRight, WristRight, ShoulderRight, ShoulderCenter, HandLeft, ShoulderLeft, WristLeft, ElbowLeft, HipRight, HipLeft, Spine, HipCenter
474	HandRight, WristRight, HandLeft, ElbowRight, WristLeft, ShoulderRight, ShoulderCenter, Head, ShoulderLeft, Spine, ElbowLeft
595	ElbowRight, HandRight, WristRight, ShoulderRight, FootRight, AnkleRight, Head, ShoulderCenter, KneeRight, HandLeft, WristLeft, Spine, HipRight, HipCenter, ShoulderLeft
644	HandRight, WristRight, ElbowRight
723	HandRight, HandLeft, WristRight, WristLeft, ElbowRight, ElbowLeft
	Touch your toes
	Adult
565	Head, HandRight, WristRight, ElbowRight, ShoulderRight, ShoulderCenter, HandLeft, WristLeft, HipRight, HipLeft, Spine, ShoulderLeft, HipCenter, ElbowLeft, KneeLeft
577	Head, ShoulderRight, ShoulderCenter, ShoulderLeft, ElbowRight, ElbowLeft, WristRight, WristLeft, HandRight, HandLeft, HipLeft, Spine, HipRight, HipCenter, KneeLeft, KneeRight

Table F-1. Continued

ID	Joints Selected
604	Head, HandRight, WristRight, ElbowRight, ShoulderRight, ShoulderCenter, HipRight, HandLeft, HipLeft, WristLeft, ShoulderLeft, Spine, HipCenter, ElbowLeft, KneeRight, KneeLeft, AnkleRight
734	Head, ShoulderRight, ShoulderLeft, ElbowLeft, ShoulderCenter, ElbowRight, WristRight, WristLeft, HandRight, HandLeft, HipRight, HipLeft, HipCenter, Spine, KneeRight, KneeLeft
876	Head, WristLeft, ShoulderRight, ElbowLeft, HandLeft, ShoulderCenter, ElbowRight, ShoulderLeft, WristRight, HandRight, HipLeft, HipCenter, HipRight, Spine, KneeLeft, KneeRight
888	Head, ShoulderLeft, ShoulderCenter, ShoulderRight, ElbowRight, ElbowLeft, WristRight, WristLeft, HandRight, HipRight, HandLeft, HipLeft, Spine, HipCenter, KneeRight, KneeLeft
921	Head, ShoulderRight, ElbowRight, WristRight, ShoulderLeft, ElbowLeft, HandRight, WristLeft, HandLeft, ShoulderCenter, Spine, HipCenter, HipRight, HipLeft, KneeRight, KneeLeft, AnkleLeft
934	Head, WristRight, HandRight, ElbowRight, ShoulderRight, ShoulderCenter, HandLeft, HipRight, ElbowLeft, WristLeft, ShoulderLeft, HipLeft, Spine, HipCenter, KneeRight, KneeLeft, AnkleRight, FootRight
970	Head, HandRight, WristRight, ShoulderCenter, ElbowRight, HandLeft, WristLeft, ShoulderRight, HipCenter, Spine, HipLeft, HipRight, ElbowLeft, ShoulderLeft, KneeRight, KneeLeft, AnkleLeft, FootLeft
976	Head, WristLeft, ShoulderCenter, ShoulderLeft, ElbowLeft, HandLeft, WristRight, ShoulderRight, HandRight, ElbowRight, HipLeft, Spine, HipCenter, HipRight, KneeLeft, KneeRight, FootLeft, AnkleLeft
	Child
103	Head, ShoulderCenter, ShoulderLeft, ShoulderRight, ElbowLeft, WristLeft, HandLeft, ElbowRight, Spine, HandRight, WristRight, HipLeft, HipRight, HipCenter
106	Head, ShoulderCenter, ShoulderRight, HandLeft, ShoulderLeft, WristLeft, HandRight, WristRight, ElbowLeft, Spine, ElbowRight, HipLeft, HipCenter, HipRight, KneeLeft, KneeRight
169	Head, ShoulderCenter, ShoulderLeft, ShoulderRight, Spine, HipCenter, HipLeft, HipRight
290	HandLeft, Head, WristLeft, HandRight, WristRight, ShoulderCenter, ElbowLeft, ShoulderLeft, ElbowRight, HipRight, HipLeft, ShoulderRight, Spine, HipCenter
337	Head, ShoulderRight, ShoulderLeft, HandRight, HandLeft, ShoulderCenter, WristLeft, ElbowLeft, WristRight, ElbowRight, Spine, HipRight, HipCenter, HipLeft, KneeRight, KneeLeft
342	HandLeft, Head, WristLeft, ShoulderLeft, ElbowLeft, HipRight, ShoulderRight, ShoulderCenter, HipCenter, Spine, HandRight, WristRight, AnkleLeft, HipLeft, KneeLeft, KneeRight, ElbowRight, FootLeft

Table F-1. Continued

ID	Joints Selected
474	ShoulderRight, Head, HandLeft, WristRight, HandRight, WristLeft, ShoulderLeft, ElbowRight, ShoulderCenter, HipRight, ElbowLeft, HipLeft, AnkleLeft, FootLeft, Spine, HipCenter
595	Head, HandLeft, WristLeft, ShoulderCenter, ElbowLeft, HandRight, WristRight, ElbowRight, ShoulderRight, ShoulderLeft, Spine, HipCenter, HipRight, FootRight, HipLeft, KneeRight
644	Head, HandRight, HandLeft, ShoulderCenter, WristRight, WristLeft, ShoulderRight, ElbowRight, ShoulderLeft, HipRight, HipLeft, ElbowLeft, Spine, HipCenter, KneeRight
723	ShoulderCenter, Head, ShoulderRight, ShoulderLeft, Spine, HipRight, HipCenter, HipLeft, HandLeft, WristLeft

APPENDIX G
CODEBOOK

Table G-1. Codebook for the qualitative analysis of children's motions

Code	Definition	Values	Example	Notes
PARTICIPANT DETAILS				
Participant ID	Fill in the ID of the participant		e.g., 290	ID is the name of the video
Age	Fill in the Age of the participant		Age corresponding to ID as in the Kinder-Gator paper	
Gender	Fill in the Gender of the participant			Gender corresponding to ID in the Kinder-Gator paper
Handedness	Fill in the dominant hand of participant			Handedness corresponding to Handedness in Kinder-Gator paper
Motion_ Type	Fill in the type of motion being performed		e.g., Raise your hand	Choose from one of the motion types specified
Side_ Used	Fill in which side of the body is raised or put forward. If upper body, the dominant should be focused on the arms. If lower body, the dominant should be focused on the legs	Left, Right, Both, Def, NA		Only for upper and lower body motions. Enter NA for full body motions. Def means no arm lifted for upper body or no leg lifted for lower body
Motion_ Category	Fill in the body part that is predominantly required to move when performing this motion	Upper Body, Lower Body, Full Body		Use the pre-defined table to fill in (A full body motion is considered as both an upper body and lower body motion)
Video_ Timestamp	Timestamp from the video when the participant performs the motion			

Table G-1. Continued

Code	Definition	Values	Example	Notes
How well do users move body parts relative to each other to produce movement				
UpperBody_ Moves_ LowerBody	Does the upper body move when a lower body motion is being performed?	Yes, No, NA (Not Applicable)	e.g., Arm raised when bending knees	If yes, then write in the body part being used. Write NA if Motion_Category is Upper Body. Put in notes if it is just the head that moves [codename: bodypart]
LowerBody_ Moves_ UpperBody	Does the lower body move when an upper body motion is being performed?	Yes, No, NA (Not Applicable)	e.g., Leg moves when throwing ball	If yes, then write in the body part being used. Write NA if Motion_Category is Lower Body
#_ Arm_ Movement	Number of arms moved to perform the motion	0 (No arm movement), 1 (only one arm), 2 (Both arms are used)	e.g., Arms raised when bending to bow	Count arm movement even if it seems unrelated to motion (e.g., scratching head)
#_ Leg_ Movement	Number of legs moved to perform the motion	0 (No leg movement), 1 (only one leg), 2 (Both legs are used)	e.g., number of legs moved to bend knee	Count leg movement even if it seems unrelated to motion
How well do users move a single body part?				
LeftArm_ Stance_ Fixed	Was the left arm moved without interruptions (i.e., at some point during the movement of a body part, did the participant stop the movement and return to default position without completing the motion or did the participant stop the movement voluntarily or involuntarily and then continue along the same direction)	Continuous, Interrupted, DEF (default: used when body part doesn't move)	e.g., raising arm but then going back to the side without completing the movement	Use default if left arm doesn't move

Table G-1. Continued

Code	Definition	Values	Example	Notes
RightArm_ Stance_ Fixed	Was the right arm moved without interruptions (i.e., at some point during the movement of a body part, did the participant stop the movement and return to default position without completing the motion or did the participant stop the movement voluntarily or involuntarily and then continue along the same direction)	Continuous, Interrupted, DEF (default)	e.g., raising arm but then going back to the side without completing the movement	Use default if right arm doesn't move/when body part is in default state (stand state)
LeftLeg_ Stance_ Fixed	Was the left leg moved without interruptions (i.e., at some point during the movement of a body part, did the participant stop the movement and return to default position without completing the motion or did the participant stop the movement voluntarily or involuntarily and then continue along the same direction)	Continuous, Interrupted, DEF (default)	e.g., Leg continues to drop on the ground when lifting leg to one side	Use default if left leg doesn't move/when body part is in default state (stand state)
RightLeg_ Stance_ Fixed	Was the right leg moved without interruptions (i.e., at some point during the movement of a body part, did the participant stop the movement and	Continuous, Interrupted, DEF (default)	e.g., Leg continues to drop on the ground when lifting leg to one side	Use default if right leg doesn't move/when body part is in default state (stand state)

Table G-1. Continued

Code	Definition	Values	Example	Notes
	return to default position without completing the motion or did the participant stop the movement voluntarily or involuntarily and then continue along the same direction)			
#_ Arm_ Bent	Quantifies the number of arms that are bent to perform the motion	0 (No arms bent), 1 (1 arm bent), 2 (Both arms bent)	e.g., One hand bent when making a bow will be 1	
#_ Leg_ Bent	Quantifies the number of legs that are bent to perform the motion	0 (No legs bent), 1 (1 leg bent), 2 (Both leg bent)	e.g., Only bending one leg during forward lunge will be 1	
Dir_ UpperBody_ Start	Predominant direction of upper body during movement toward the peak pose	Along-XY, Along-XZ, Along-YZ, DEF	e.g., putting hands on hips and leaning to side can be characterized as Along-XY	Put in notes if it is just the head that moves [codename: bodypart] Peak: pose that signifies the actual performance of the motion (last pose before participant starts to return to default stance). Default State: Initial standing pose wherein the participant is not moving. Use default if upper body doesn't move/when body part is in default state (stand state).
Dir_ UpperBody_ Peak	Direction of upper body at the peak pose	Up (vertically upwards; 90 degrees up) [U],	e.g., putting hands on hips and leaning	Put in notes if it is just the head that moves

Table G-1. Continued

Code	Definition	Values	Example	Notes
		Down (vertically downwards; 90 degrees down) [D], Forward (Toward Camera; perpendicular to the body) [F], Forward-Side [FS], Forward-Up [FU], Forward-Down [FD], Backward(Away from Camera; perpendicular to the body) [B], Side (horizontally straight; 90 degrees side way) [S], Side-Down [SD], Side-Up [SU], Default [DEF]	to side can be characterized as S.	[codename: bodypart]. Peak: pose that signifies the actual performance of the motion (last pose before participant starts to return to default stance). Default State: Initial standing pose wherein the participant is not moving. Use default if upper body doesn't move/when body part is in default state (stand state)
Dir_UpperBody_End	Predominant direction of upper body from after the peak pose back to the default state	Along-XY, Along-XZ, Along-YZ, DEF	e.g., putting hands on hips and leaning to side can be characterized as Along-XY	Put in notes if it is just the head that moves [codename: bodypart]. Peak: pose that signifies the actual performance of the motion (last pose before participant starts to return to default stance). Default State: Initial standing pose wherein the participant is not moving. Use default if upper body doesn't move/when body part is in default state (stand state)

Table G-1. Continued

Code	Definition	Values	Example	Notes
Dir_ Left-Arm_ Start	Predominant direction of left arm during movement toward the peak pose	-XY, Along-XZ Along-YZ, DEF	e.g., Arm raised high when raising arm to one side can be characterized as Along-XY	Put in notes if it is just the head that moves [codename: bodypart]. Peak: pose that signifies the actual performance of the motion. Default State: Initial standing pose wherein the participant is not moving. Use default if upper body doesn't move/when body part is in default state (stand state). Left: Is toward the fridge. Right: Is toward the dispenser
Dir_ Left-Arm_ Peak	Direction of left arm at the peak pose	Up (vertically upwards; 90 degrees up) [U], Down (vertically downwards; 90 degrees down) [D], Forward (Toward Camera; perpendicular to the body) [F], Forward-Side [FS], Forward-Up [FU], Forward-Down [FD], Backward(Away from Camera; perpendicular to the body) [B], Side (horizontally straight; 90 degrees side ways)	e.g., Arm raised high when raising arm to one side can be characterized as S	Put in notes if it is just the head that moves [codename: bodypart]. Peak: pose that signifies the actual performance of the motion (last pose before participant starts to return to default stance). Default State: Initial standing pose wherein the participant is not moving. Use default if upper body doesn't move/when body part is in default state (stand state). Left: Is toward the fridge. Right: Is toward the dispenser

Table G-1. Continued

Code	Definition	Values	Example	Notes
		[S], Side-Down [SD], Side-Up [SU], Default [DEF]		
Dir_ Left-Arm_ End	Predominant direction of left arm from after the peak pose back to the default state	Along-XY, Along-XZ Along-YZ, DEF	e.g., Arm raised high when raising arm to one side can be characterized as Along-XY	Put in notes if it is just the head that moves [codename: bodypart]. Peak: pose that signifies the actual performance of the motion (last pose before participant starts to return to default stance). Default State: Initial standing pose wherein the participant is not moving. Use default if upper body doesn't move/when body part is in default state (stand state). Left: Is toward the fridge. Right: Is toward the dispenser
Dir_ RightArm_ Start	Predominant direction of right arm during movement toward the peak pose	Along-XY, Along-XZ, Along-YZ, DEF	e.g., Arm raised high when raising arm to one side can be characterized as Along-XY	Put in notes if it is just the head that moves [codename: bodypart]. Peak: pose that signifies the actual performance of the motion (last pose before participant starts to return to default stance). Default State: Initial standing pose wherein the participant is not moving. Use default if upper body doesn't move/when body part is in default state (stand state).

Table G-1. Continued

Code	Definition	Values	Example	Notes
				Left: Is toward the fridge. Right: Is toward the dispenser
Dir_ RightArm_ Peak	Direction of right arm at the peak pose	Up (vertically upwards; 90 degrees up) [U], Down (vertically downwards; 90 degrees down) [D], Forward (Toward Camera; perpendicular to the body) [F], Forward-Side [FS], Forward-Up [FU], Forward-Down [FD], Backward(Away from Camera; perpendicular to the body) [B], Side (horizontally straight; 90 degrees side way) [S], Side-Down [SD], Side-Up [SU], Default [DEF]	e.g., Arm raised high when raising arm to one side can be characterized as SU	Put in notes if it is just the head that moves [codename: bodypart]. Peak: pose that signifies the actual performance of the motion (last pose before participant starts to return to default stance). Default State: Initial standing pose wherein the participant is not moving. Use default if upper body doesn't move/when body part is in default state (stand state). Left: Is toward the fridge. Right: Is toward the dispenser
Dir_ RightArm_ End	Predominant direction of right arm from after the peak pose back to the default state	Along-XY, Along-XZ, Along-YZ, DEF	e.g., Arm raised high when raising arm to one side can be characterized as Along-XY	Put in notes if it is just the head that moves [codename: bodypart]. Peak: pose that signifies the actual performance of the motion (last pose before participant starts to return to default stance). Default State: Initial

Table G-1. Continued

Code	Definition	Values	Example	Notes
				standing pose wherein the participant is not moving. Use default if upper body doesn't move/when body part is in default state (stand state). Left: Is toward the fridge. Right: Is toward the dispenser
Dir_ Left-Knee_ Start	Predominant direction of left knee during movement toward the peak pose	-XY, Along-XZ, Along-YZ, DEF	e.g., Lift leg to one side can be characterized as Along-XY	Put in notes if it is just the head that moves [codename: bodypart]. Peak: pose that signifies the actual performance of the motion (last pose before participant starts to return to default stance). Default State: Initial standing pose wherein the participant is not moving.Use default if upper body doesn't move/when body part is in default state (stand state). Left: Is toward the fridge. Right: Is toward the dispenser
Dir_ Left-Knee_ Peak	Direction of left knee at the peak pose	Up (vertically upwards; 90 degrees up) [U], Down (vertically downwards; 90 degrees down) [D], Forward (Toward Camera; perpendicular to the body) [F], Forward-Side	e.g., Lift leg to one side can be characterized as S	Put in notes if it is just the head that moves [codename: bodypart]. Peak: pose that signifies the actual performance of the motion (last pose before partici-

Table G-1. Continued

Code	Definition	Values	Example	Notes
		[FS], Forward-Up [FU], Forward-Down [FD], Backward(Away from Camera; perpendicular to the body) [B], Side (horizontally straight; 90 degrees side way) [S], Side-Down [SD], Side-Up [SU], Default [DEF]		pant starts to return to default stance). Default State: Initial standing pose wherein the participant is not moving. Use default if upper body doesn't move/when body part is in default state (stand state). Left: Is toward the fridge. Right: Is toward the dispenser.
Dir_ Left-Knee_ End	Predominant direction of left knee from after the peak pose back to the default state	Along-XY, Along-XZ, Along-YZ, DEF	e.g., Lift leg to one side can be characterized as Along-XY	Put in notes if it is just the head that moves [codename: bodypart]. Peak: pose that signifies the actual performance of the motion (last pose before participant starts to return to default stance). Default State: Initial standing pose wherein the participant is not moving. Use default if upper body doesn't move/when body part is in default state (stand state).
Dir_ RightKnee_ Start	Predominant direction of right knee during movement toward the peak pose	Along-XY, Along-XZ, Along-YZ, DEF	e.g., Lift leg to one side can be characterized as Along-XY	Put in notes if it is just the head that moves [codename: bodypart]. Peak: pose that signifies the actual performance of the motion (last pose before participant starts to return to default stance).

Table G-1. Continued

Code	Definition	Values	Example	Notes
				Default State: Initial standing pose wherein the participant is not moving. Use default if upper body doesn't move/when body part is in default state (stand state). Left: Is toward the fridge. Right: Is toward the dispenser
Dir_ RightKnee_ Peak	Direction of right knee at the peak pose	Up (vertically upwards; 90 degrees up) [U],Down (vertically downwards; 90 degrees down) [D], Forward (Toward Camera; perpendicular to the body) [F], Forward-Side [FS], Forward-Up [FU], Forward-Down [FD], Backward(Away from Camera; perpendicular to the body) [B], Side (horizontally straight; 90 degrees side way) [S], Side-Down [SD], Side-Up [SU],Default [DEF]	e.g., Lift leg to one side can be characterized as S	Put in notes if it is just the head that moves [codename: bodypart]. Peak: pose that signifies the actual performance of the motion (last pose before participant starts to return to default stance). Default State: Initial standing pose wherein the participant is not moving. Use default if upper body doesn't move/when body part is in default state (stand state). Left: Is toward the fridge. Right: Is toward the dispenser.
Dir_ RightKnee_ End	Predominant direction of right knee from after the peak pose back to the	Along-XY, Along-XZ, Along-YZ, DEF	e.g., Lift leg to one side can be characterized	Put in notes if it is just the head that moves [codename: bodypart]. Peak:

Table G-1. Continued

Code	Definition	Values	Example	Notes
	default state.		as Along-XY	pose that signifies the actual performance of the motion (last pose before participant starts to return to default stance). Default State: Initial standing pose wherein the participant is not moving. Use default if upper body doesn't move/when body part is in default state (stand state). Left: Is toward the fridge. Right: Is toward the dispenser.
Dir_ LeftLeg_ Start	Predominant direction of entire left leg during movement toward the peak pose	Along-XY, Along-XZ, Along-YZ, DEF	e.g., Lift leg to one side can be characterized as Along-XY	Put in notes if it is just the head that moves [codename: bodypart]. Peak: pose that signifies the actual performance of the motion (last pose before participant starts to return to default stance). Default State: Initial standing pose wherein the participant is not moving. Use default if upper body doesn't move/when body part is in default state (stand state). Left: Is toward the fridge. Right: Is toward the dispenser.
Dir_ LeftLeg_ Peak	Direction of entire left leg at the peak pose	Up (vertically upwards; 90 degrees up) [U], Down (vertically	e.g., Lift leg to one side can be characterized as S	Put in notes if it is just the head that moves [codename: bodypart]. Peak:

Table G-1. Continued

Code	Definition	Values	Example	Notes
		downwards; 90 degrees down) [D], Forward (Toward Camera; perpendicular to the body) [F], Forward-Side [FS], Forward-Up [FU], Forward-Down [FD], Backward(Away from Camera; perpendicular to the body) [B], Side (horizontally straight; 90 degrees side way) [S], Side-Down [SD], Side-Up [SU], Default [DEF]		pose that signifies the actual performance of the motion (last pose before participant starts to return to default stance). Default State: Initial standing pose wherein the participant is not moving. Use default if upper body doesn't move/when body part is in default state (stand state). Left: Is toward the fridge.Right: Is toward the dispenser.
Dir_LeftLeg_End	Predominant direction of entire left leg from after the peak pose back to the default state	Along-XY, Along-XZ, Along-YZ, DEF	e.g., Lift leg to one side can be characterized as Along-XY	Put in notes if it is just the head that moves [codename: bodypart]. Peak: pose that signifies the actual performance of the motion (last pose before participant starts to return to default stance). Default State: Initial standing pose wherein the participant is not moving. Use default if upper body doesn't move/when body part is in default state (stand state). Left: Is toward the fridge. Right: Is toward the

Table G-1. Continued

Code	Definition	Values	Example	Notes
				dispenser.
Dir_ Right-Leg_ Start	Predominant direction of entire right leg during movement toward the peak pose	Along-XY, Along-XZ, Along-YZ, DEF	e.g., Lift leg to one side can be characterized as Along-XY	Put in notes if it is just the head that moves [codename: bodypart]. Peak: pose that signifies the actual performance of the motion (last pose before participant starts to return to default stance). Default State: Initial standing pose wherein the participant is not moving. Use default if upper body doesn't move/when body part is in default state (stand state). Left: Is toward the fridge. Right: Is toward the dispenser.
Dir_ Right-Leg_ Peak	Direction of entire right leg at the peak pose	Up (vertically upwards; 90 degrees up) [U], Down (vertically downwards; 90 degrees down) [D], Forward (Toward Camera; perpendicular to the body) [F], Forward-Side [FS], Forward-Up [FU], Forward-Down [FD], Backward(Away from Camera; perpendicular to the body) [B], Side	e.g., Lift leg to one side can be characterized as S	Put in notes if it is just the head that moves [codename: bodypart]. Peak: pose that signifies the actual performance of the motion (last pose before participant starts to return to default stance). Default State: Initial standing pose wherein the participant is not moving. Use default if upper body doesn't move/when body part is in default state (stand state). Left: Is toward the fridge. Right: Is toward the dispenser.

Table G-1. Continued

Code	Definition	Values	Example	Notes
		(horizontally straight; 90 degrees side way) [S], Side-Down [SD], Side-Up [SU], Default [DEF]		
Dir_ Right-Leg_ End	Predominant direction of entire right leg from after the peak pose back to the default state	Along-XY, Along-XZ, Along-YZ, DEF	e.g., Lift leg to one side can be characterized as Along-XY	Put in notes if it is just the head that moves [codename: bodypart]. Peak: pose that signifies the actual performance of the motion (last pose before participant starts to return to default stance). Default State: Initial standing pose wherein the participant is not moving. Use default if upper body doesn't move/when body part is in default state (stand state). Left: Is toward the fridge. Right: Is toward the dispenser
How forceful is the motion performed				
Effort_ UpperBody	Effort applied to the upper body to perform the motion	Light (less effort than referent), normal (same effort as referent), extra(more effort than referent), DEF (Default state)	e.g., slight bent of upper body to make bow	Referents located at onedrive (kindergator/ RGB Videos/ Referents)

Table G-1. Continued

Code	Definition	Values	Example	Notes
Effort_ Arm	Effort applied to the arm to perform the motion	Light (less effort than referent), normal (same effort as referent), extra(more effort than referent), DEF (Default state)	e.g., moving arm far back to punch	Referents located at onedrive (kinder-gator/ RGB Videos/ Referents)
Effort_ Leg	Effort applied to the leg to perform the motion	Light (less effort than referent), normal (same effort as referent), extra(more effort than referent), DEF (Default state)	e.g., deep bend of knees to jump	Referents located at onedrive (kinder-gator/ RGB Videos/ Referents)
How fast is the motion performed				
Body_ Speed	How fast did the user move their body when performing the motion?	Slow (less speed than referent), Normal (same speed as referent), Fast (more speed than referent).	e.g., swinging arms to swipe	Referents located at onedrive (kinder-gator/RGB Videos/Referents)

REFERENCES

[1] Jake K. Aggarwal and Michael S. Ryoo. 2011. Human activity analysis: A review. *Comput. Surveys* 43, 3 (2011), 1–43. https://doi.org/10.1145/1922649.1922653

[2] Justin Amadeus Albert, Victor Owolabi, Arnd Gebel, Clemens Markus Brahms, Urs Granacher, and Bert Arnrich. 2020. Evaluation of the pose tracking performance of the azure kinect and kinect v2 for gait analysis in comparison with a gold standard: A pilot study. *Sensors* 20, 18 (2020), 5104.

[3] Aishat Aloba, Gianne Flores, Julia Woodward, Alex Shaw, Amanda Castonguay, Isabella Cuba, Yuzhu Dong, Eakta Jain, and Lisa Anthony. 2018. Kinder-Gator: The UF kinect database of child and adult motion.. In *Eurographics (Short Papers)*, Olga Diamanti and Amir Vaxman (Eds.). 13–16. https://doi.org/10.2312/egs.20181033

[4] Aishat Aloba, Annie Luc, Julia Woodward, Yuzhu Dong, Rong Zhang, Eakta Jain, and Lisa Anthony. 2019. Quantifying differences between child and adult motion based on gait features. In *International Conference on Human-Computer Interaction (HCII '19)*. Springer, 385–402. https://doi.org/10.1007/978-3-030-23563-5_31

[5] Aishat Aloba, Julia Woodward, and Lisa Anthony. 2020. FilterJoint: Toward an Understanding of Whole-Body Gesture Articulation. In *International Conference on Multimodal Interaction (ICMI '20)*. Springer, 213–221. https://doi.org/10.1145/3382507.3418822

[6] Lisa Anthony, Quincy Brown, Jaye Nias, Berthel Tate, and Shreya Mohan. 2012. Interaction and recognition challenges in interpreting children's touch and gesture input on mobile devices. In *Proceedings of the ACM Conference on Interactive Tabletops and Surfaces (ITS '12)*. 225–234. https://doi.org/10.1145/2396636.2396671

[7] Lisa Anthony, Radu-Daniel Vatavu, and Jacob O. Wobbrock. 2013. Understanding the consistency of users' pen and finger stroke gesture articulation. In *Proceedings of the Graphics Interface Conference (GI '13)*. 87–94. https://doi.org/10.5555/2532129.2532145

[8] Lisa Anthony and Jacob O. Wobbrock. 2010. A Lightweight Multistroke Recognizer for User Interface Prototypes. *Proceedings of Graphics Interface (GI '10)*, 245–252. https://doi.org/10.5555/1839214.1839258

[9] Lisa Anthony and Jacob O. Wobbrock. 2012. $N-protractor: a fast and accurate multistroke recognizer. In *Proceedings of Graphics Interface (GI '12)*. 117–120. https://doi.org/10.5555/2305276.2305296

[10] Tiago V. Barreira, Peter T. Katzmarzyk, William D. Johnson, and Catrine Tudor-Locke. 2013. Walking cadence and cardiovascular risk in children and adolescents: NHANES, 2005-2006. *American Journal of Preventive Medicine* 45, 6 (2013). https://doi.org/10.1016/j.amepre.2013.08.005

[11] Catherine E. Bauby and Arthur D. Kuo. 2000. Active control of lateral balance in human walking. *Journal of Biomechanics* 33, 11 (2000), 1433–1440. https://doi.org/10.1016/S0021-9290(00)00101-9

[12] Tony Belpaeme, Paul E. Baxter, Robin Read, Rachel Wood, Heriberto Cuayáhuitl, Bernd Kiefer, Stefania Racioppa, Ivana Kruijff-Korbayová, Georgios Athanasopoulos, Valentin Enescu, Rosemarijn Looije, Mark Neerincx, Yiannis Demiris, Raquel Ros-Espinoza, Aryel Beck, Lola Cañamero, Antione Hiolle, Matthew Lewis, Ilaria Baroni, Marco Nalin, Piero Cosi, Giulio Paci, Fabio Tesser, Giacomo Sommavilla, and Remi Humbert. [n.d.]. ([n. d.]).

[13] MG Benedetti and A Cappozzo. 1994. Anatomical Landmark Definition and Identification in Computer Aided Movement Analysis in a Rehabilitation Context II (Internal Report). *U Degli Studi La Sapienza* (1994), 1–31.

[14] Bir Bhanu and Ju Han. 2002. Bayesian-based performance prediction for gait recognition. In *Workshop on Motion and Video Computing (MOTION '02)*. 145–150. https://doi.org/10.1109/MOTION.2002.1182227

[15] Leslie Bishko. 2014. Animation principles and laban movement analysis - movement frameworks for creatinc empathic character performances. *Nonverbal Communication in Virtual Worlds: Understanding and Designing Expressive Characters* (2014), 177–203.

[16] Aaron F. Bobick. 1997. Movement, activity and action: The role of knowledge in the perception of motion. *Philosophical Transactions of the Royal Society B: Biological Sciences* 352, 1358 (1997), 1257–1265. https://doi.org/10.1098/rstb.1997.0108

[17] Aaron F. Bobick and James W. Davis. 2001. The recognition of human movement using temporal templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2001), 257–267. https://doi.org/10.1109/34.910878

[18] William Bruce. 1910. Exercise in Education and Medicine. *The Lancet* 175, 4521 (1910), 1165–1166.

[19] Sait Celebi, Ali S. Aydin, Talha T. Temiz, and Tarik Arici. 2013. Gesture recognition using skeleton data with weighted dynamic time warping. In *International Conference on Computer Vision Theory and Applications (VISAPP '13)*. https://doi.org/10.5220/0004217606200625

[20] Chen Chen, Roozbeh Jafari, and Nasser Kehtarnavaz. 2015. UTD-MHAD: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor. In *Proceedings of the International Conference on Image Processing (ICIP '15)*. 168–172. https://doi.org/10.1109/ICIP.2015.7350781

[21] Belkacem Chikhaoui, Bing Ye, and Alex Mihailidis. 2017. Feature-level combination of skeleton joints and body parts for accurate aggressive and agitated behavior recognition. *Journal of Ambient Intelligence and Humanized Computing* 8, 6 (2017), 957–976.

[22] Enea Cippitelli, Samuele Gasparrini, Ennio Gambi, and Susanna Spinsante. 2016. A Human Activity Recognition System Using Skeleton Data from RGBD Sensors. *Computational Intelligence and Neuroscience* (2016), 14 pp. https://doi.org/10.1155/2016/4351435

[23] Frances Cleland-Donnelly, Suzanne S Mueller, and David L Gallahue. 2017. *Developmental Physical Education for All Children: Theory Into Practice*. Vol. 5.

[24] William T. Cochran, James W. Cooley, David L. Favin, Howard D. Helms, Reginald A. Kaenel, William W. Lang, George C. Maling, David E. Nelson, Charles M. Rader, and Peter D. Welch. 1967. What is the Fast Fourier Transform? *IEEE Transactions on Audio and Electroacoustics* 55, 10 (1967), 1664–1674.

[25] Rory A. Cooper. 1995. *Rehabilitation Engineering Applied to Mobility and Manipulation*. Vol. 1. CRC Press. 516 pages.

[26] Corinna Cortes and Vladimir Vapnik. 1995. Support-Vector Networks. *Machine Learning* (1995). https://doi.org/10.1023/A:1022627411411

[27] F. Danion, E. Varraine, M. Bonnard, and J. Pailhous. 2003. Stride variability in human gait: the effect of stride frequency and stride length. *Gait & Posture* 18, 1 (2003), 69–77. https://doi.org/10.1016/S0966-6362(03)00030-4

[28] James W Davis. 2001. Visual categorization of children and adult walking styles. In *International Conference on Audio-and Video-based Biometric Person Authentication*. Springer, 295–300. https://doi.org/10.1007/3-540-45344-X_43

[29] Debapratim Das Dawn and Soharab Hossain Shaikh. 2016. A comprehensive survey of human action recognition with spatio-temporal interest point (STIP) detector. *The Visual Computer* 32, 3 (2016), 289–306.

[30] Curriculum Development and Supplemental Materials Commission. [n.d.]. *Physical education framework for California public schools, kindergarten through grade twelve*.

[31] Yuzhu Dong, Aishat Aloba, Sachin Paryani, Lisa Anthony, Neha Rana, and Eakta Jain. 2017. Adult2Child: Dynamic scaling laws to create child-like motion. In *SIGGRAPH Conference on Motion, Interaction, and Games (MIG '17)*. 1–10. https://doi.org/10.1145/3136457.3136460

[32] Janto F. Dreijer and Ben M. Herbst. 2008. Action Classification using the Average of Pose Changes. In *Symposium of the Pattern Recognition Association of South Africa*. 85–85.

[33] Liqiang Du, Hong Chen, Shuli Mei, and Qing Wang. 2016. Real-time human action recognition using individual body part locations and local joints structure. In *ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry (VRCAI '16)*. 293–298. https://doi.org/10.1145/3013971.3013974

[34] Carlos Duarte, Daniel Costa, Luis Carriço, André Falcão, David Costa, and Luís Tavares. 2014. Welcoming gesture recognition into autism therapy. In *SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. https://doi.org/10.1145/2559206.2581337

[35] Stacey C Dusing and Deborah E Thorpe. 2007. A normative sample of temporal and spatial gait parameters in children using the GAITRite(R) electronic walkway. *Gait & Posture* 25, 1 (2007), 135–139.

[36] Deanna Fish and Jean-Paul Nielsen. 1993. Clinical Assessment of Human Gait. *Journal of Prosthetics and Orthotics* 5, 2 (1993), 39–48.

[37] Simon Fothergill, Helena M. Mentis, Pushmeet Kohli, and Sebastian Nowozin. 2012. Instructing people for training gestural interactive systems. In *SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. https://doi.org/10.1145/2207676.2208303

[38] Robert Fox and Cynthia McDaniel. 1982. The perception of biological motion by human infants. *Science* 218, 4571 (oct 1982), 486–487. https://doi.org/10.1126/science.7123249

[39] James R. Gage, Peter A. Deluca, and Thomas S. Renshaw. 1995. Gait Analysis: Principles and Applications. *The Journal of Bone & Joint Surgery* 77, 10 (1995), 1607 – 1623. http://jbjs.org/content/77/10/1607.abstract

[40] Guillermo Garcia-Hernando and Tae-Kyun Kim. 2017. Transition forests: Learning discriminative temporal transitions for action recognition and detection. In *IEEE Conference on Computer Vision and Pattern Recognition*. 432–440.

[41] Arnold Gesell, Frances L. Ilg, Louise B. Ames, and Glenna E. Bullis. 1946. *The child from five to ten*. Harper & Brothers.

[42] Elena Gianaria, Marco Grangetto, Maurizio Lucenteforte, and Nello Balossino. 2014. Human classification using gait features. In *International Workshop on Biometric Authentication*, Vol. 8897. 16–27. https://doi.org/10.1007/978-3-319-13386-7_2

[43] Gene H. Golub and Christian Reinsch. 1970. Singular value decomposition and least squares solutions. *Numer. Math.* (1970). https://doi.org/10.1007/BF02163027

[44] Jacqueline D. Goodway, John C. Ozmun, and David L. Gallahue. 2002. *Understanding motor development: Infants, children, adolescents, adults*.

[45] Shannon E. Gray and Caroline F. Finch. 2015. The causes of injuries sustained at fitness facilities presenting to Victorian emergency departments - identifying the main culprits. *Injury Epidemiology* 2, 1 (2015), 1–8.

[46] Gutemberg Guerra-Filho and Arnab Biswas. 2012. The human motion database: A cognitive and parametric sampling of human motion. *Image and Vision Computing* 30, 3 (2012), 251–261.

[47] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. 2002. Gene selection for cancer classification using support vector machines. *Machine learning* 46, 1 (2002), 389–422.

[48] Tibor Guzsvinecz, Veronika Szucs, and Cecilia Sik-Lanyi. 2019. Suitability of the Kinect sensor and Leap Motion controller—A literature review. *Sensors* 19, 5 (2019), 1072.

[49] Kilem L Gwet. 2014. *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters*. Advanced Analytics, LLC.

[50] Tomasz Hachaj and Marek R. Ogiela. 2014. Rule-Based Approach to Recognizing Human Body Poses and Gestures in Real Time. *Multimedia Systems* 20, 1 (2014), 81–99.

[51] Kevin A Hallgren. 2012. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology* 8, 1 (2012), 23.

[52] Aran Hampapur, Lisa Brown, Jonathan Connell, Sharat Pankanti, Andrew Senior, and Yingli Tian. 2003. Smart surveillance: Applications, technologies and implications. In *Joint Conference of the International Conference on Information, Communications and Signal Processing and Pacific-Rim Conference on Multimedia (ICICS-PCM '03)*. 1133–1138. https://doi.org/10.1109/ICICS.2003.1292637

[53] John A. Hartigan and Manchek A. Wong. 2006. Algorithm AS 136: A K-Means clustering algorithm. *Journal of the Royal Statistical Society Series C (Applied Statistics)* 29, 1 (2006), 100–108.

[54] Kathleen M Haywood and Nancy Getchell. 2019. *Life span motor development*. Human kinetics.

[55] Mathew W. Hill, Maximilian M. Wdowski, Adam Pennell, David F. Stodden, and Michael J. Duncan. 2019. Dynamic postural control in children: Do the arms lend the legs a helping hand? *Frontiers in Physiology* 113 (2019), 272–280. https://doi.org/10.3389/fphys.2018.01932

[56] Zaher Hinbarji, Rami Albatal, and Cathal Gurrin. 2015. Dynamic user authentication based on mouse movements curves. In *International Conference on Multimedia Modeling (MMM '15)*. 111–122. https://doi.org/10.1007/978-3-319-14442-9_10

[57] Johanna Hoysniemi. 2006. International survey on the Dance Dance Revolution game. *Computers in Entertainment* 4, 2 (2006), 8–8. https://doi.org/10.1145/1129006.1129019 arXiv:05/1000 [1544-3574]

[58] Weiming Hu, Dan Xie, Zhouyu Fu, Wenrong Zeng, and Steve Maybank. 2007. Semantic-based surveillance video retrieval. *IEEE Transactions on Image Processing* 16, 4 (2007), 1168–1171.

[59] Donald F. Huelke. 1998. An Overview of Anatomical Considerations of Infants and Children in the Adult World of Automobile Safety Design. *Association for the Advancement of Automotive Medicine* 42 (1998), 93–113. https://doi.org/10.1145/982452.982461 arXiv:arXiv:1011.1669v3

[60] Joseph P. Hunter, Robert N. Marshall, and Peter J. McNair. 2004. Interaction of Step Length and Step Rate during Sprint Running. *Medicine and Science in Sports and Exercise* 36, 2 (2004), 261–271. https://doi.org/10.1249/01.MSS.0000113664.15777.53

[61] Mohamed E Hussein, Marwan Torki, Mohammad A Gowayyed, and Motaz El-Saban. 2013. Human action recognition using a temporal hierarchy of covariance descriptors on 3d joint locations. In *International joint conference on artificial intelligence*.

[62] Pham Chinh Huu, Le Quoc Khanh, and Le Thanh Ha. 2014. Human Action Recognition Using Dynamic Time Warping and Voting Algorithm. *VNU Journal of Science: Comp. Science & Com. Eng* 30, 3 (2014), 22–30.

[63] Bon Woo Hwang, Sungmin Kim, and Seong Whan Lee. 2006. A full-body gesture database for automatic gesture recognition. In *International Conference on Automatic Face and Gesture Recognition (FG '06)*. 243–248. https://doi.org/10.1109/FGR.2006.8

[64] Eakta Jain, Lisa Anthony, Aishat Aloba, Amanda Castonguay, Isabella Cuba, Alex Shaw, and Julia Woodward. 2016. Is the Motion of a Child Perceivably Different from the Motion of an Adult? *ACM Transactions on Applied Perception* 13, 4 (jul 2016), 1–17. https://doi.org/10.1145/2947616

[65] Harshad Kadu, Maychen Kuo, and C. C. Jay Kuo. 2011. Human motion classification and management based on mocap data analysis. In *ACM Multimedia Conference and Co-Located Workshops (MM '11)*. 73–74. https://doi.org/10.1145/2072572.2072594

[66] Geetanjali Vinayak Kale and Varsha Hemant Patil. 2016. A study of vision based human motion recognition and analysis. *International Journal of Ambient Computing and Intelligence* (2016). https://doi.org/10.4018/IJACI.2016070104 arXiv:1608.06761

[67] Justin J. Kavanagh, R. S. Barrett, and Steven Morrison. 2004. Upper body accelerations during walking in healthy young and elderly men. *Gait and Posture* 20, 3 (2004), 291–298. https://doi.org/10.1016/j.gaitpost.2003.10.004

[68] Yu Kong and Yun Fu. 2018. Human Action Recognition and Prediction: A Survey. (2018).

[69] Sven Kratz and Michael Rohs. 2010. A $3 gesture recognizer: simple gesture recognition for devices equipped with 3D acceleration sensors. In *Proceeding of the international conference on Intelligent user interfaces (IUI '10)*. 341–344. https://doi.org/10.1145/1719970.1720026

[70] Sven Kratz and Michael Rohs. 2011. Protractor3d: a closed-form solution to rotation-invariant 3d gestures. In *Proceedings of the International Conference on Intelligent User Interfaces (IUI '11)*. 371–374. https://doi.org/10.1145/1943403.1943468

[71] Gordon Kurtenbach and Eric A. Hulteen. 1990. Gestures in human-computer communications. In *The art of human computer interface design*, Laurel Brenda (Ed.). Addison-Wesley, Boston, MA, 309–317.

[72] Rudolf Laban and Lisa Ullmann. 1971. The mastery of movement. (1971).

[73] J. Richard Landis and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.

[74] Pierre Le Bras, David A. Robb, Thomas S. Methven, Stefano Padilla, and Mike J. Chantler. 2018. Improving user confidence in concept maps: Exploring data driven explanations. In *SIGCHI Conference on Human Factors in Computing Systems (CHI '18)*. 1–13.

[75] Wanqing Li, Zhengyou Zhang, and Zicheng Liu. 2010. Action recognition based on a bag of 3D points. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops (CVPRW '10)*. 9–14. https://doi.org/10.1109/CVPRW.2010.5543273

[76] Yang Li. 2010. Protractor: a fast and accurate gesture recognizer. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. 2169–2172. https://doi.org/10.1145/1753326.1753654

[77] Hongyi Liu and Lihui Wang. 2018. Gesture recognition for human-robot collaboration: A review. *International Journal of Industrial Ergonomics* 68 (2018), 355–367.

[78] Roanna Lun and Wenbing Zhao. 2015. A Survey of Applications and Human Motion Recognition with Microsoft Kinect. *International Journal of Pattern Recognition and Artificial Intelligence* 29, 5 (2015), 1555008–1555008.

[79] Roanna Lun and Wenbing Zhao. 2015. A Survey of Applications and Human Motion Recognition with Microsoft Kinect. *International Journal of Pattern Recognition and Artificial Intelligence* January (2015), 49. https://doi.org/10.1142/S0218001415550083

[80] Andrew Macvean and Judy Robertson. 2012. iFitQuest: a school based study of a mobile location-aware exergame for adolescents. *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '12)* (2012), 359. https://doi.org/10.1145/2371574.2371630

[81] Andrew Macvean and Judy Robertson. 2013. Understanding Exergame Users' Physical Activity, Motivation and Behavior over Time. *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*, 1251–1260. https://doi.org/10.1145/2470654.2466163

[82] Mentar Mahmudi and Marcelo Kallmann. 2015. Multi-modal data-driven motion planning and synthesis. In *ACM SIGGRAPH Conference on Motion in Games (MIG '15)*. 119–124.

[83] Ilias El Makrini, Kelly Merckaert, Dirk Lefeber, and Bram Vanderborght. 2017. Design of a collaborative architecture for human-robot assembly tasks. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '17)*. 1624–1629. https://doi.org/10.1109/IROS.2017.8205971

[84] Sebastián Maldonado and Richard Weber. 2009. A wrapper method for feature selection using support vector machines. *Information Sciences* 179, 13 (2009), 2208–2217.

[85] Gaëlle Marais and Patrick Pelayo. 2003. Cadence and exercise: physiological and biomechanical determinants of optimal cadences–practical applications. *Sports Biomech* 2, 1 (Jan 2003), 103–132.

[86] M Martínez-Zarzuela, FJ Díaz-Pernas, A Tejeros-de Pablos, D González-Ortega, and M Antón-Rodríguez. 2014. Action recognition system based on human body tracking with depth images. *Advances in Computer Science: an International Journal* 3, 1 (2014), 115–123.

[87] Grant McCracken. 1988. *The long interview*. Vol. 13. Sage.

[88] Jiangyuan Mei, Meizhu Liu, Yuan Fang Wang, and Huijun Gao. 2016. Learning a Mahalanobis Distance-Based Dynamic Time Warping Measure for Multivariate Time Series Classification. *IEEE Transactions on Cybernetics* 26, 6 (2016), 1363–1374. https://doi.org/10.1109/TCYB.2015.2426723

[89] Microsoft. 2019. Kinect for Windows. https://developer.microsoft.com/en-us/windows/kinect

[90] Microsoft. 2021. Azure Kinect DK. https://azure.microsoft.com/en-us/services/kinect-dk/

[91] Rolf Moe-Nilssen and Jorunn L. Helbostad. 2004. Estimation of gait cycle characteristics by trunk accelerometry. *Journal of Biomechanics* 37, 1 (2004), 121–126. https://doi.org/10.1016/S0021-9290(03)00233-1

[92] Charles Molnar and Jane Gair. 2013. Chapter 19.3: Joints and Skeletal Movement. In *Concepts of Biology-1st Canadian Edition*.

[93] Samiul Monir, Sabirat Rubya, and Hasan Shahid Ferdous. 2012. Rotation and scale invariant posture recognition using microsoft kinect skeletal tracking feature. In *International Conference on Intelligent Systems Design and Applications (ISDA '12)*. 404–409. https://doi.org/10.1109/ISDA.2012.6416572

[94] GS Don Morris. 1980. *Elementary School Physical Education: Toward Inclusion*. Brighton Publishing Company.

[95] Sarah Morrison-Smith, Aishat Aloba, Hangwei Lu, Brett Benda, Shaghayegh Esmaeili, Gianne Flores, Jesse Smith, Nikita Soni, Isaac Wang, Rejin Joy, Woodard Damon, Jaime Ruiz, and Lisa Anthony. 2020. MMGatorAuth: A Novel Multimodal Dataset for Authentication Interactions in Gesture and Voice. In *Proceedings of the 2020 International Conference on Multimodal Interaction*. 370–377. https://doi.org/10.1145/3382507.3418881

[96] Meinard Müller, Tido Röder, Michael Clausen, Bernhard Eberhardt, Björn Krüger, and Andreas Weber. 2007. Documentation Mocap Database HDM05. *Database* (2007), 34. https://doi.org/10.1016/0260-4779(90)90032-9

[97] Lewis M. Nashner. 1980. Balance adjustments of humans perturbed while walking. *Journal of neurophysiology* 44, 4 (1980), 650–664. https://doi.org/10.1152/jn.1980.44.4.650

[98] Jasmir Nijhar, Nadia Bianchi-Berthouze, and Gemma Boguslawski. 2012. Does Movement Recognition Precision affect the Player Experience in Exertion Games?. In *International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN '12)*. 73–82. https://doi.org/10.1007/978-3-642-30214-5_9

[99] Shahriar Nirjon, Chris Greenwood, Carlos Torres, Stefanie Zhou, John A. Stankovic, Hee Jung Yoon, Ho Kyeong Ra, Can Basaran, Taejoon Park, and Sang H. Son. 2014. Kintense: A robust, accurate, real-time and evolving system for detecting aggressive actions from streaming 3D skeleton data. In *IEEE International Conference on Pervasive Computing and Communications (PerCom '14)*. 2–10. https://doi.org/10.1109/PerCom.2014.6813937

[100] Juliet Norton, Chadwick A. Wingrave, and Joseph J. LaViola. 2010. Exploring strategies and guidelines for developing full body video game interfaces. In *International Conference on the Foundations of Digital Games (FDG '10)*. 155–162. https://doi.org/10.1145/1822348.1822369

[101] Tommy Öberg, Alek Karsznia, and Kurt Öberg. 1993. Basic gait parameters: reference data for normal subjects, 10-79 years of age. *Journal of rehabilitation research and development* 30 (1993), 210–210.

[102] Ferda Ofli, Rizwan Chaudhry, Gregorij Kurillo, René Vidal, and Ruzena Bajcsy. 2014. Sequence of the most informative joints (SMIJ): A new representation for human skeletal action recognition. *Journal of Visual Communication and Image Representation* 25 (2014).

[103] Yi Chung Pai and James Patton. [n.d.]. ([n. d.]).

[104] Georgios Th Papadopoulos, Apostolos Axenopoulos, and Petros Daras. 2014. Real-time skeleton-tracking-based human action recognition using kinect data. In *International Conference on Multimedia Modeling (MMM '14)*. Springer, 473–483.

[105] Orasa Patsadu, Chakarida Nukoolkit, and Bunthit Watanapa. 2012. Human gesture recognition using Kinect camera. In *International Joint Conference on Computer Science and Software Engineering (JCSSE '12)*. 28–32. https://doi.org/10.1109/JCSSE.2012.6261920

[106] V. Gregory Payne and Larry D. Isaacs. 2011. *Human Motor Development: A LifeSpan Approach*. Routledge. https://doi.org/10.4324/9781315213040

[107] Leif Peterson. 2009. K-nearest neighbor. *Scholarpedia* (2009), 1883–1883. https://doi.org/10.4249/scholarpedia.1883

[108] Jean Piaget and Susan Wedgwood. 2015. *The grasp of consciousness: Action and concept in the young child*. Psychology Press. https://doi.org/10.4324/9781315722382

[109] Rare and Ultimate Play the Game. 2010. Kinect Sports.

[110] Santiago Riofrío, David Pozo, Jorge Rosero, and Juan Vásques. 2017. Gesture Recognition using Dynamic Time Warping and Kinect: A Practical Approach. In *International Conference on Information Systems and Computer Science (ICICT '17)*. 302–308.

[111] Mikel D. Rodriguez, Javed Ahmed, and Mubarak Shah. 2008. Action MACH: A Spatio-Temporal Maximum Average Correlation Height Filter for Action Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*. 6–6. https://doi.org/10.1109/CVPR.2008.4587727

[112] Viviana Rota, Laura Perucca, Anna Simone, and Luigi Tesio. [n.d.]. ([n. d.]).

[113] Heydar Sadeghi, Paul Allard, François Prince, and Hubert Labelle. 2000. Symmetry and limb dominance in able-bodied gait: A review. , 34–45 pages. https://doi.org/10.1016/S0966-6362(00)00070-9

[114] John W Santrock. 1995. *Life-span development*. WCB Brown & Benchmark Publishers.

[115] Christian Schüldt, Ivan Laptev, and Barbara Caputo. 2004. Recognizing Human Actions: A Local SVM Approach. In *International Conference on Pattern Recognition (ICPR '04)*. 32–36. https://doi.org/10.1109/ICPR.2004.1334462

[116] Lorenzo Seidenari, Vincenzo Varano, Stefano Berretti, Alberto Del Bimbo, and Pietro Pala. 2013. Recognizing actions from depth cameras as weakly aligned multi-part bag-of-poses. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '13)*. 479–485. https://doi.org/10.1109/CVPRW.2013.77

[117] Rachael D. Seidler, Jessica A. Bernard, Taritonye B. Burutolu, Brett W. Fling, Mark T. Gordon, Joseph T. Gwin, Youngbin Kwak, and David B. Lipps. 2010. (2010).

[118] Noboru Sekiya and Hiroshi Nagasaki. 1998. Reproducibility of the walking patterns of normal young adults: test-retest reliability of the walk ratio(step-length/step-rate). *Gait & Posture* 7, 3 (1998), 225–227.

[119] R C. Shaefer. 1987. Body Alignment, Postures, and Gait. In *Clinical biomechanics: musculoskeletal actions and reactions*. Wiliams & Wilkins. https://www.chiro.org/ACAPress/Body{_}Alignment.html

[120] Alex Shaw. 2020. *Automatic Recognition of Children's Touchscreen Stroke Gestures*. Ph.D. Dissertation. University of Florida.

[121] Alex Shaw and Lisa Anthony. 2016. Analyzing the Articulation Features of Children's Touchscreen Gestures. In *ACM International Conference on Multimodal Interaction (ICMI '16)*. 333–340. https://doi.org/10.1145/2993148.2993179

[122] Sarah P. Shultz, Andrew P. Hills, Michael R. Sitler, and Howard J. Hillstrom. 2010. Body size and walking cadence affect lower extremity joint power in children's gait. *Gait & posture* 32, 2 (2010), 248–252.

[123] Rawesak Tanawongsuwan and Aaron Bobick. 2003. A Study of Human Gaits across Different Speeds. *Georgia Technology* (2003), 1–13.

[124] Tiffany Y. Tang and Relic Yongfu Wang. 2016. A comparative study of applying low-latency smoothing filters in a multi-kinect virtual play environment. In *International Conference on Human Computer Interaction (HCII '16)*. 144–148. https://doi.org/10.1007/978-3-319-40542-1_23

[125] Pattreeya Tanisaro, Florian Mahner, and Gunther Heidemann. 2017. Quasi view-independent human motion recognition in subspaces. In *International Conference on Machine Learning and Computing (ICMLC '17)*. 278–283. https://doi.org/10.1145/3055635.3056577

[126] Weijun Tao, Tao Liu, Rencheng Zheng, and Hutian Feng. 2012. Gait analysis using wearable sensors. *Sensors* 12, 2 (2012), 2255–2283.

[127] Eugene M. Tarnata, Amirreza Samiei, Mehran Maghoumi, Pooya Khaloo, Corey R. Pittman, and Joseph J. LaViola. 2017. Jackknife: A reliable recognizer with few samples and many modalities. In *SIGCHI Conference on Human Factors in Computing Systems (CHI '17)*. 5850–5861. https://doi.org/10.1145/3025453.3026002

[128] Phillippe Terrier. 2012. Step-to-Step Variability in Treadmill Walking: Influence of Rhythmic Auditory Cueing. *PLoS One* 7, 10 (2012), e47171. https://doi.org/10.1371/journal.pone.0047171

[129] Michal Tölgyessy, Martin Dekan, L'uboš Chovanec, and Peter Hubinskỳ. 2021. Evaluation of the Azure Kinect and Its Comparison to Kinect V1 and Kinect V2. *Sensors* 21, 2 (2021), 413.

[130] Catrine Tudor-Locke, Sarah M. Camhi, Claudia Leonardi, William D. Johnson, Peter T. Katzmarzyk, Conrad P. Earnest, and Timothy S. Church. 2011. Patterns of adult stepping cadence in the 2005-2006 NHANES. *Preventive Medicine* 53, 3 (2011), 178–181. https://doi.org/10.1016/j.ypmed.2011.06.004

[131] Ubisoft. 2017. Just Dance 2018.

[132] Radu-Daniel Vatavu. 2017. Beyond features for recognition: human-readable measures to understand users' whole-body gesture performance. *International Journal of Human–Computer Interaction* 33, 9 (2017), 713–730.

[133] Radu Daniel Vatavu. 2017. Improving gesture recognition accuracy on touch screens for users with low vision. In *SIGCHI Conference on Human Factors in Computing Systems (CHI '17)*. 4667–4679. https://doi.org/10.1145/3025453.3025941

[134] Radu-Daniel Vatavu, Lisa Anthony, and Jacob O Wobbrock. 2012. Gestures as point clouds: A $P recognizer for user interface prototypes. In *ACM International Conference on Multimedia Interaction (ICMI '12)*. 273–280. https://doi.org/10.1145/2388676.2388732

[135] Radu-Daniel Vatavu, Lisa Anthony, and Jacob O. Wobbrock. 2013. Relative accuracy measures for stroke gestures. In *International Conference on Multimodal Interaction (ICMI '13)*. 279–286. https://doi.org/10.1145/2522848.2522875

[136] Aliza Vigderman. 2020. . https://www.security.org/smart-home/consumer-adoption-2020/

[137] Federico Visi, Rodrigo Schramm, Esther Coorevits, and Eduardo Reck Miranda. 2017. Musical instruments, body movement, space, and motion data: Music as an emergent multimodal choreography. *Human Technology* 13, 1 (2017), 58–81.

[138] Ekaterina Volkova, Stephan De La Rosa, Heinrich H. Bülthoff, and Betty Mohler. 2014. The MPI emotional body expressions database for narrative scenarios. *PLoS ONE* 9, 12 (2014), e113647.

[139] Isaac Wang, Mohtadi Ben Fraj, Pradyumna Narayana, Dhruva Patil, Gururaj Mulay, Rahul Bangar, J. Ross Beveridge, Bruce A. Draper, and Jaime Ruiz. 2017. EGGNOG: A continuous, multi-modal data set of naturally occurring gestures with ground truth labels. In *IEEE International Conference on Automatic Face Gesture Recognition (FG '17)*. 414–421.

[140] Qifei Wang, Gregorij Kurillo, Ferda Ofli, and Ruzena Bajcsy. 2015. Evaluation of pose tracking accuracy in the first and second generations of microsoft kinect. In *International conference on healthcare informatics (ICHI '15)*. IEEE, 380–389. https://doi.org/10.1109/ICHI.2015.54.

[141] Diane Watson, Regan L. Mandryk, and Kevin G. Stanley. 2013. The Design and Evaluation of a Classroom Exergame. In *Proceedings of the International Conference on Gameful Design, Research, and Applications (Gamification '13)*. 34–41. https://doi.org/10.1145/2583008.2583013

[142] Wilhelm Weber and Eduard Weber. 1992. *Mechanics of Human Walking Apparatus*. Springer-Verlag.

[143] Wei Wei and Yunxiao An. 2009. Vision-based human motion recognition: A survey. In *International Conference on Intelligent Networks and Intelligent Systems (ICINIS '09)*. 386–389. https://doi.org/10.1109/ICINIS.2009.105

[144] Mark Weiser. 1999. The computer for the 21st century. *ACM SIGMOBILE mobile computing and communications review* 3, 3 (1999), 3–11.

[145] David A. Winter. 1995. Human balance and posture control during standing and walking. *Gait and Posture* 3, 4 (1995), 193–214. https://doi.org/10.1016/0966-6362(96)82849-9

[146] Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 143–146.

[147] Jacob O. Wobbrock, Andrew D. Wilson, and Yang Li. 2007. Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. In *ACM symposium on User interface software and technology (UIST '07)*. 159–168. https://doi.org/10.1145/1294211.1294238

[148] Tzu Tsung Wong. 2015. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition* 48, 9 (2015), 2839–2846.

[149] Julia Woodward, Alex Shaw, Aishat Aloba, Ayushi Jain, Jaime Ruiz, and Lisa Anthony. 2017. Tablets, tabletops, and smartphones: cross-platform comparisons of children's touchscreen interactions. In *International Conference on Multimodal Interaction (ICMI '17)*. ACM Press, New York, New York, USA, 5–14. https://doi.org/10.1145/3136755.3136762

[150] Julia Woodward, Alex Shaw, Annie Luc, Brittany Craig, Juthika Das, Phillip Hall Jr, Akshay Holla, Germaine Irwin, Danielle Sikich, Quincy Brown, et al. 2016. Characterizing how interface complexity affects children's touchscreen interactions. In *SIGCHI Conference on Human Factors in Computing Systems (CHI '16)*. 1921–1933.

[151] Adam Woznica, Phong Nguyen, and Alexandros Kalousis. 2012. Model mining for robust feature selection. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '12)*. 913–921. https://doi.org/10.1145/2339530.2339674

[152] Qingqiang Wu, Guanghua Xu, Longting Chen, Ailing Luo, and Sicong Zhang. 2017. Human action recognition based on kinematic similarity in real time. *Plos one* 12, 10 (2017), e0185719.

[153] Lu Xia, Chia Chih Chen, and J. K. Aggarwal. 2012. View invariant human action recognition using histograms of 3D joints. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW '12)*. 20–27. https://doi.org/10.1109/CVPRW.2012.6239233

[154] Jinghua Yu, Qing Wang, and Hong Chen. 2018. Application of kinect-based motion recognition algorithm in cultural tourism. In *International Conference on Network, Communication and Computing (ICNCC '18)*. 307–311. https://doi.org/10.1145/3301326.3301377

[155] Liwei Zhao and Norman I Badler. 2001. *Synthesis and acquisition of laban movement analysis qualitative parameters for communicative gestures*. Institute for Research in Cognitive Science Technical Report IRCS-01-11. University of Pennysylvania.

[156] Dongsheng Zhou, Xinzhu Feng, Pengfei Yi, Xin Yang, Qiang Zhang, Xiaopeng Wei, and Deyun Yang. 2019. 3D human motion synthesis based on convolutional neural network. *IEEE Access* 7 (2019), 66325–66335.

BIOGRAPHICAL SKETCH

Aishat Aloba received her B.SC in computer science with a minor in mathematics from the University of Mississippi in May 2015. She began her graduate program at the University of Florida that same year and received her M.SC in computer science in December 2019 and her Ph.D in Human-centered computing in August 2021. Her research interests include Human-Computer Interaction, Child-Computer Interaction, Multimodal Interaction, and Motion-Based Games.