© 2019 by Roshni Kaushik. All rights reserved.

DEVELOPING AND EVALUATING A MODEL FOR HUMAN MOTION TO FACILITATE LOW DEGREE-OF-FREEDOM ROBOT IMITATION OF HUMAN MOVEMENT

BY

ROSHNI KAUSHIK

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Mechanical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2019

Urbana, Illinois

Advisor:

Assistant Professor Amy LaViers

Abstract

Imitation of human motion is a necessary activity for robots to integrate seamlessly into human-facing environments. While perfect replication is not possible, especially for low degree-of-freedom (DOF) robots, this thesis presents a model for human motion that achieves perceptual imitation. Motion capture data of dyadic interactions was first analyzed to quantify a characteristic of human motion observed in the movement. The leaning of the spine, or verticality, was found to correlate with these movement observations. Verticality was then used to inspire a low-DOF model of human motion using motion capture that can be used to command the movement of simulated robots. Experiments were developed to test users' perception of the imitation by these 3 and 4-DOF simulated robots of human motion. Verticality was preferred in an initial study over artificially generated motion for the higher DOF robot, Broombot, which was preferred over the lower DOF robot, Rollbot. A study was developed to test the preferences of users when the mapping between human and robot motion was changed for variable human motion. Motion capture-based motion was preferred over artificially generated motion, and a sub-group of respondents who preferred verticality and were more engaged in the survey was found. Since the experiments were performed using motion capture data from a trained ballet dancer, a discussion of the differences between two Indian classical dance styles is included that shows that verticality alone is not representative of all motion and prompts a further analysis to develop socially adaptive robot behavior. In-progress and future work include a hardware implementation that will allow real-time motion capture data to drive simulated and/or physical robots. *Menagerie* is an in-development performance using the tools developed in this thesis that can include a human with simulated and/or physical robots moving together.

Acknowledgments

First and foremost, I want to thank Amy, without whom this work would not have been possible. From our first phone conversation to our many meetings, you have always guided me in the best way. You always pushed me to produce my best work, helping me gain invaluable research skills and shaping me into the student and person I am today.

I also want to thank my collaborators and lab colleagues who inspired me and troubleshooted with me to bring this from just an inkling of an idea to this entire thesis. Umer, thank you for your support and friendship throughout this process, and we had so many great conversations about politics and religion and sometimes even research. Erin, your help to record motion capture data for this thesis was invaluable. Ilya and Cat, your input early on in this work provided much needed insight to form the foundation of this thesis. I also want to thank the rest of the RAD Lab for always being supportive and providing much needed feedback throughout my work. I am also thankful to Kathy in the MechSe Administrative office and Allison in the MechSe Business office for always being available for my academic and travel questions.

And of course to my parents and brother, who survived my phone calls and chaotic travel.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

How should two bodies of different morphology move to imitate one another? While exact replication of activity is not possible, particularly when one system has many fewer degrees-of-freedom than the other, we can define *perceptual imitation* to be achieved when human viewers see the same activity between the two bodies. Even between humans, our differing mobility and limb lengths create differences in the execution of a task. Yet, even with these differences, we imitate each other in many situations including children learning new movements from the adults around them, people in exercise classes following the movements of an instructor, or pedestrians taking cues from the people around them on the sidewalk. Moreover, simple cartoon characters and robots are frequently seen as "doing the same thing" as natural counterparts. These examples show that this perceptual imitation is possible and common.

One definition of imitation from sociology limits it to behaviors fulfilling (1) the imitated behavior being new for the imitator, (2) the same task strategy of the demonstrator being employed, and (3) the same task goal being accomplished [4]. Focusing on the third definition, we will look at whether and how simple, simulated robots can imitate human movement. A survey of robot imitation of humans in [5] details the utility of imitation as a mode of robot movement including the creation of a shared lexicon between two movers and discusses the open problem of determining which part of human movement is 'relevant' to imitate. A survey of socially interactive robots similarly discusses the questions of how the robot knows what to imitate and how it maps observed action into behavior [6].

In this thesis, we explore the relationships between human motion, robot motion, and the perception of imitation. This chapter reviews relevant background information and related work in the areas of dyadic interactions (Section 1.1), virtual characters (Section 1.2), and robotics (Section 1.3). Chapter 2 explores the development of a measure tracking the leaning of the spine, verticality, to characterize motion capture examples of dyadic motion. Chapter 3 creates a model for human imitation by low degree-of-freedom artificial systems inspired by this verticality measure. In Chapter 4, we consider the imitation effectiveness of verticality on two low-DOF systems for human imitation and analyze the changing perception of imitation when the human motion and human-robot mapping are varied. Chapter 5 analyzes human motion of two

Indian classical dance styles and how the observed motion can be transferred to robots. Chapter 6 concludes and details in-progress and future work.

1.1 Dyadic interactions

A number of studies have been conducted on dyadic interactions, or interactions between pairs of individuals. For example, analyzing the making and breaking of symmetry of the head (mirror symmetry) during conversations was shown to be a meaningful element of communication when modeled with a dynamical system [7, 8, 9]. The pattern of this imitation influenced how highly participants rated conversation quality. Head motion specifically, during human conversation (nods, tilts, etc.), was recorded for robots to display more "natural" nonverbal cues during a human interaction [10].

A co-manipulation task of a dyad moving a table, characterized by the forces applied by their hands, was used to explore correlations between the various task parameters and the the performance of the pair, including comparisons to minimum jerk trajectories [11]. An analysis of the handshake using sensor data from many users decomposed the features of the motion into phases [12]. Both of these prior efforts have worked to characterize human motion for the development of robotic counterparts.

Optical motion capture tracks a set of reflective markers on a human over time during a movement, while active sensors can also be used to directly measure acceleration of body parts. Both produce a low-dimensional skeleton that reduces the complex, high degree-of-freedom system of the human body to a finite set of typically 30 rotating and translating points in 3-D space. Although optical capture does have its limitations such as how "natural" the movements recorded are [13], it remains one of the best ways to capture detailed, three-dimensional records of human motion.

For example, active motion capture has been used to quantify key features in tango dancing [14], measuring the bending of dancers' chests to identify individual performance and correlating between individual postures to characterize collaborative performance. This metric is similar to one of our mappings based on the leaning of the spine, first proposed in [15]. Similar active capture tools, which prevent occlusion during data collection, were used to develop an interactive application, or tool, that facilitated feedback between two dancers and their musical accompaniment. Limb-based metrics were used to determine movement similarity, but motion features were not classified [16].

Optical motion capture was used to categorize movement in terms of valence and arousal parameters to classify motion by emotion [17]. Dance performances have been synthesized by combining a musical analysis and a motion analysis using motion capture that utilized intensity and the key frames on the music

beats [18]. Motion preferences of humans were calculated and analyzed by users comparing the movements of various amoebas to a motion capture bhangra dancer to determine which amoeba behaviors mimicked the dancer's movement best; this work used a pairwise comparison of mappings in order to create a task with low cognitive effort for human subjects [19]. Researchers have asked users to imitate a video of arm movements and recorded data using sensors similar to motion capture, and the imitation was evaluated using a joint-space based segmentation and comparison algorithm [20]. These methods are examples of a large body of work proposing different *mappings* of motion capture data between bodies to measure or quantify imitation.

Joint improvisation, or the creative action of two or more people without a script or designated leader, was studied in the context of imitation using a mirror game and motion capture, and a model for the observed behavior was generated [21, 22]. Machine learning and neural networks can be used to abstract away the complexities of interaction by training models with examples. Gaussian Mixture Models (GMM) of Interaction Primitives model nonlinear correlations between different movers [23, 24]. An interactive online dance work explores how different movement representations can elicit different perceptions from an audience, or kinesthetic empathy [25]. An evolutionary dynamics approach using replicator-mutator dynamics has been used to model dominance – achieved through encouraging behavior switching between pre-defined modes of unison – in group dance performances [26]. Thus, the notion of "moving together" is often clarified through the lens of imitation, which allows for individual-specific differences in action while finding similarity.

1.2 Virtual Characters

Virtual characters offer greater ability for imitation of human behavior than physical systems because they are not limited by actuator and control system performance. These simulated systems also afford the opportunity for complex skeletons, which similarly do not require advanced hardware development for motion generation. In this thesis, we use simulated characters, without motor performance limits, that have only a few degrees of freedom, making our work more portable to robotic systems (discussed in the next section).

One mapping between human and virtual character poses was generated by using a few key human poses from a motion capture session of a human moving in a way similar to the desired movement of the character [27]. A methodology for creating correspondences between groups of body parts on different characters allowed for the generation of motions on different characters from the same source [28]. In a puppetry context, motion for animated non-human characters was generated using direct feature mapping

using dynamical equations before a human motion-capture movement was used to control the characters in real-time [29]. These examples again show the diversity of proposed mappings between human and artificial motion.

Using a Kinect sensor to capture a skeleton, a platform was developed for human-virtual character interactions for a variety of applications [30]. A dancing game was developed that generated a real-time animation of virtual dancers on a screen that corresponded to the motion capture movement of human dancers using a matching algorithm that recognized specific dance moves [31]. Joint improvisation has also been used in the realm of virtual character development. Researchers designed a mirror game in which a human was able to successfully improvise during a mirror game, a framework for imitation, with a virtual agent [32, 33]. One thing that imitation provides to a human interactant is the sense that the artificial system is aware of their motion, building a kinesthetic interaction channel, which imitating robots might also leverage.

Machine learning techniques have also been used to generate movement of virtual characters, under the guiding lens of large datasets. Researchers used learning from demonstration to program a virtual dancer by developing an internal model of a human dancer's movements using Artificial Neural Networks (ANN) and Hidden Markov Models (HMM) and reacting to some movements from a human dancer [34]. A Gaussian Mixture Model (GMM) trained with examples of two humans interacting recognized new actions and generated responses of a virtual character [35].

1.3 Robotics

In the field of robotics, like the area of virtual characters, many methods have been developed for mapping human movement to robot platforms. Mapping human motion to robot motion presents many difficulties including measurement and motion feasibility [36]. A humanoid robot has been controlled by virtually connecting human motion capture markers to points on the robot with translational springs in order to find correspondences [37]. Additionally, a mapping between human and humanoid robot markers was investigated in order to emulate human behaviors during social interaction [38]. Key frame selection and a cluster-based framework was used to imitate and evaluate human motion test the imitation of children with autism with a robot [39]. Each of these prior works shows the many possible mappings that exist between recorded human motion and robotic platforms, even with humanoid target platforms.

Essential postures of the arms and step primitives in the legs were extracted from human motion capture data to generate motion on a humanoid robot to imitate the original human movements [40]. A 'simulation

theory of the mind' approach was used in an experiment with one robot imitating the actions of another through a distributed system of inverse and forward models [41]. Artificial Neural Networks (ANN) were used to generate movement on a humanoid robot that imitates human arm movements from video and marker-tracking data [42].

A mapping was created between human motion capture data and a humanoid robot motion using an optimization subject to dynamic and physical constraints of the robot [43]. Mathematical mappings between dissimilar bodies have been used to evaluate imitation, acknowledging the 'observer-dependence' of the evaluation [44]. The performance of a robotic system designed for human imitation was evaluated with quantitative system-centered assessment in combination with qualitative human-centered assessment [45]. A gesture method was developed to quickly generate positions for social robots with different morphologies [46].

Much work has focused on simpler target platforms as well. Keepon robots were used in conjunction with the Laban Effort system to imitate the movements of children dancing to music [47]. Researchers used the low-DOF head motions of a Keepon and Nao robot to convey internal state of the robot, basing the movement on Laban efforts [48, 49]. A quadroter was used to communicate affect, also using the Laban efforts, through its flight path [50]. Researchers used a mobile manipulator consisting of a robotic arm mounted on a mobile base to test motion planning software [51]. An interactive sculpture was developed as a robotic system that interacted with human observers to generate continuously evolving behavior [52].

In the area of pair-wise interactions, coupled inverted pendulums can be used as a model for dancers performing the waltz, so a human mover can move with an imitating robot follower in basic steps [53]. An energy metric was developed by imitating human trajectories on wheeled robots that related to observers' assessment of the human performance in the context of salsa [54]. Haptic feedback, a way to measure the forces a user exerts on an interface, is another tool used to understand model and imitate human motion. A dancing robot adjusted the length of its stride based on haptic feedback from the physical connection between robot and human [55], and male and female partner dancing behavior was synthesized based on haptic interactions and stride length [56].

Thus, socially-aware robots need to be equipped with imitation capabilities as this behavior is important for interaction tasks as well as the development of expressive robotic counterparts. In this work, we deepen our understanding of the motion preferences for a low-DOF simulated robot, toward the end of a simple mobile robot that successfully imitates human motion.

Chapter 2

Developing Verticality through an Analysis of Dyadic Movement

Human movement is a complex physical phenomenon, full of the richness of contexts, interactions, and variations. In particular, the intricacies of dyadic movement raise many research questions, including the manner of nonverbal communication between a pair or dyad performing a task together. In something as simple as moving a table across a room, two individuals communicate through the movement of their bodies in addition to the forces applied on the table and the floor. In partner dance, this communication channel is even more nuanced. When dissecting dyadic human movement, we seek to identify properties describing and characterizing these interactions.

This chapter presents an analysis of dyadic motion from existing motion capture data using a single degree of freedom measure that represents the leaning of the spine, or verticality. Section 2.1 examines partnering from a dance perspective to motivate the choice of verticality to analyze dyadic motion. Section 2.2 characterizes the verticality measure. The results of the verticality analysis (Section 2.3) and the quantification of a measure of the *resistance* in the motion (Section 2.4) are included as well as a brief summary of this chapter in Section 2.5.

2.1 Understanding Partnering to Motivate the Selection of Verticality

Within distinct dance styles, dyadic motion seems to demand accord between partners on the appropriate conventions for negotiating movement[1]. For example, if two people are moving together within the context of a social dance party, they will likely be moving in a way that is significantly different than if they are moving together in the context of a competitive dance event. For ease of (non-verbal) communication, the partnering agents will agree about which physical cues are meaningful and what constitutes an appropriate response.

This agreement may be prescribed by following the accepted conventions of a particular movement form,

[1]This discussion in this section is included with thanks to my collaborator, Ilya Vidrin.

all of which place distinct constraints on aesthetic values (i.e. bending and straightening the legs in a waltz versus keeping them soft in a jive [57], or de-emphasizing the line of the body [58]). Decisions are made in real-time by each party based on interpreting physical actions they direct at one another. The quality with which weight is shifted, individually and in coordination with others, seems to be an integral component to the success of the expressed intention, regardless of whether the movement is extemporaneously generated or previously choreographed.

Before defining particular aesthetic values, it is clear that motion may be evaluated in two ways: kinesthetically for the performers and visually by observers. In certain cases of coordinated movement, partners must evaluate movement both visually and kinesthetically. It is interesting to note that coordinated interaction with a partner is dependent on a certain level of trust, in terms of the physical intimacy of touch and proximity, as well as the positions that are compromising physically (such as a lift).

Thus, it seems there are consequences at stake if agents are not aware of the ways they influence each other physically. This assumes, of course, that agents want to level with their partners. Explicitly misleading an agent, while beyond the scope of this thesis, is nevertheless a contentious thought when considering how our smallest actions influence and are interpreted by partners. The fact that there might be consequences within the act of moving with others opens up an ethical dimension of understanding weight.

At its simplest, this may be expressed as understanding the physical relationship between weight, anatomic structures, and trust. For example, physically sensing the position of one's pelvis in space relative to that of one's partner, including rotation (toward or away from one's own body), tilt (up or down), and the surrounding muscular activation. Outside of technical anatomy and physiology, one major question that emerges is how this subtlety can be captured and expressed?

The quality with which one resists the force of a partner can reveal valuable insight about the other's position and weight distribution, creating an opportunity to move together in more distinct and nuanced ways. While high levels of tension and resistance may limit mobility, it is less obvious that a subtle understanding of oppositional forces is often the secret to beautiful partnering. In accurately evaluating the level of resistance of one's partner (physically, through a form of feedback, and visually, through a form of feedforward), one can create more controlled patterns, including higher lifts and faster turns.

This is especially evident in the screenshots from different pairs of dancers performing the same choreography (Petite Mort). These images, captured at the same instant in the music, display clearly the effects of differing resistance in a cooperative movement. The first couple executes a supported penchée (standing split) with each dancer on their own and has lower resistance in the movement (Figure 2.1a). The second couple executes the same position, but with a visual opposition in the movement, corresponding to a higher

(a) Screenshot from a pair performing Petite Mort with lower resistance within the interaction [59]. The dancers are Roberto Bolle and Greta Hodgkinson performing at Stars of World Ballet Gala Concert, Teatro alla Scala, Milan, Italy in 2006.

(b) Screenshot from a pair performing Petite Mort with higher resistance within the interaction [60]. The dancers are Johan Inger and Elke Schepers performing at Lucent Danstheater, The Hague, Netherlands in 1996.

Figure 2.1: Examples of varying resistance between two different dyads performing the same movement from Petite Mort. The choreography is by Jiří Kylián, and the music is Wolfgang Amadeus Mozart's Piano Concerto No. 21 in C Major Andante.

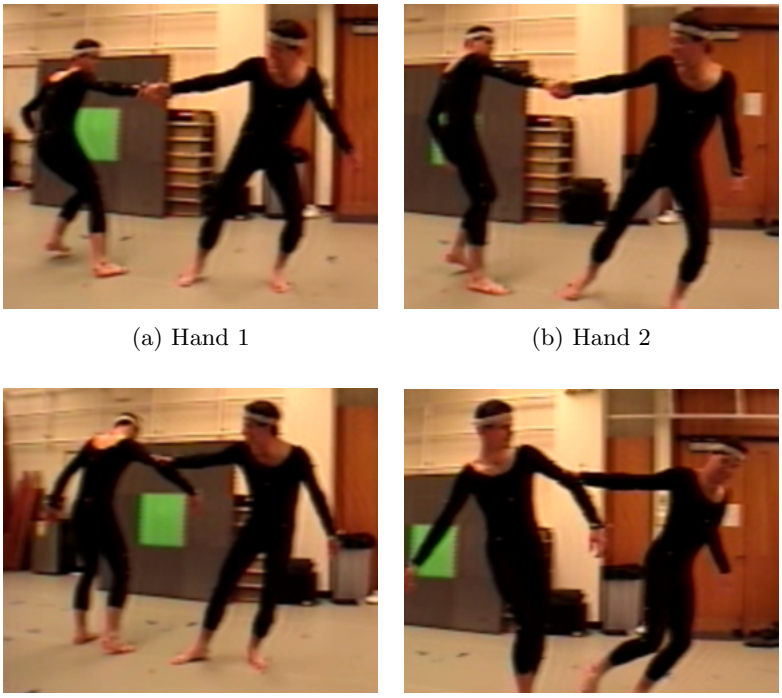
level of resistance (Figure 2.1b). The latter couple creates a different artistic expression. One could argue that the latter is a more believable partnership, given that each dancer is communicating their position to the other, as well as relying on the body of their partner to interdependently support balance and control.

We attempt to look at visual cues of basic coordinated tasks to make sense of which parameters may be at play. Clearly, the way each dancer distributes weight through their center is crucially different in these two examples. However, motion capture data of this area of the body is difficult to collect. In [61], reflective markers were surgically implanted in spinal vertebrae in order to gain some insight. Our measure will need to access broader, gross movement of the shape evolution of each partner. Thus, we look to the vertical alignment of the spine to monitor, through a low dimensional signal, bodily interactions in a dyad pair.

2.2 Description of Verticality Metric

In this section, we will describe the dataset, introduce our model from the motion capture data, and demonstrate how we calculate a one-dimensional verticality metric.

We analyzed four trials of motion capture data from the Carnegie Mellon Motion Capture Database [1]. Each of these datasets was of one person pulling another across a room, with contact point as either the hand or elbow (screenshots shown in Figure 2.2). However, we visually saw differences in the way the maneuver of pulling was executed, differences that we sought to capture quantitatively. To test our observations, we



(a) Hand 1 (b) Hand 2

(c) Elbow 1 (d) Elbow 2

Figure 2.2: Screenshots from the videos from each of the four motion capture datasets. Classified here by point of contact (hand or elbow), with two videos in each category [1].

sent the four videos to a Certified Movement Analysis (CMA) [2] to determine what an expert, qualitatively, saw as differences in the motion profiles.

The expert was instructed to comment on the four videos. Without further prompting, the expert used the word "resistance" to describe the movements in each of the videos. Table 2.1 shows excerpts from the comments we received, specifically the portions related to this concept of resistance. For Hand 2 and Elbow 2, the expert described the movement using phrases such as "less resistance" and "little resistance to being pulled." For Hand 1 and Elbow 1, she characterized the movement by "lots of resistance" and "more resistance," but for Elbow 1, the "resistance diminish[ed]" slightly over the course of the action.

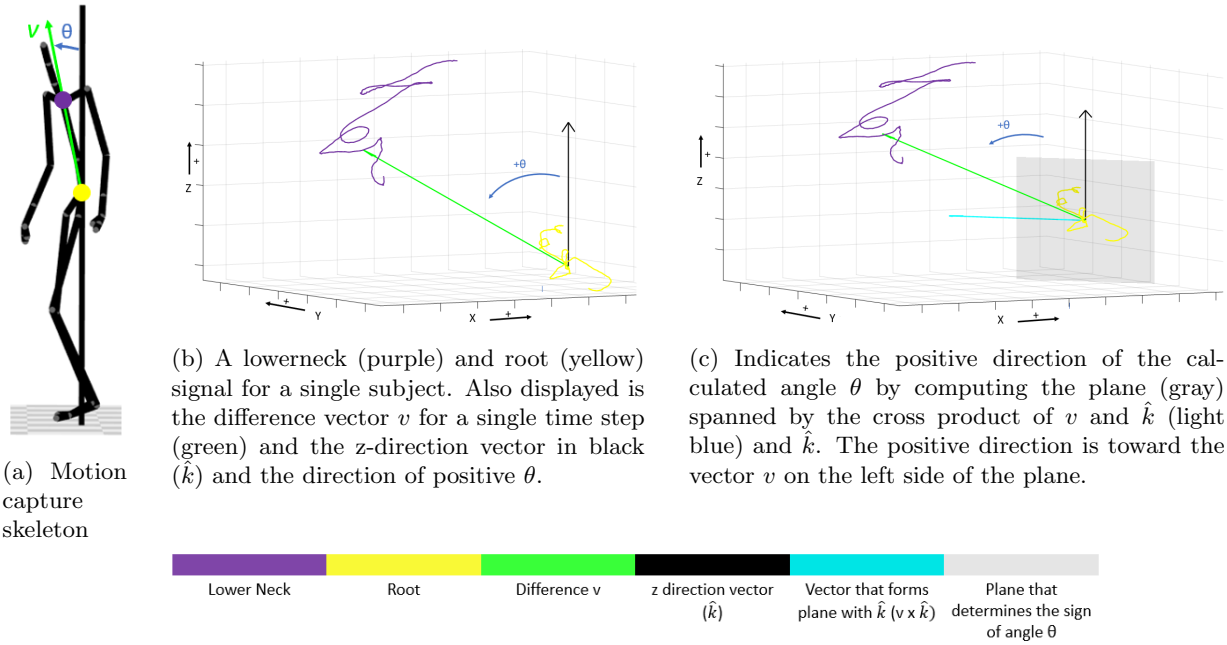
This resistance parameter that she observed qualitatively is a guiding principle for our analysis in our goal to compute quantitative metrics describing the data. With the four videos (two with hand as the point of contact and two with elbow), we separated the trials by subject. Subject A was the individual pulling, and Subject B was the individual being pulled. The two subjects across the four videos resulted in a total of 8 distinct motion profiles.

The input to our analysis was the raw motion capture data in the `.amc/.asf` format. Using MATLAB

[2]We would like to thank Catherine Maguire, CMA for her naïve expert observation of the videos of the four dyad motion samples analyzed here.

Table 2.1: Excerpts from expert comments about 4 dyadic motion videos

Video	Experts from expert comments
Hand 1	Lots of resistance to the pull
Hand 2	Less resistance to being pulled
Elbow 1	More resistance from the [person pulling] initially...but the resistance diminishes over the course of the action
Elbow 2	Little resistance to being pulled



(a) Motion capture skeleton

(b) A lowerneck (purple) and root (yellow) signal for a single subject. Also displayed is the difference vector v for a single time step (green) and the z-direction vector in black ($\hat{k}$) and the direction of positive θ.

(c) Indicates the positive direction of the calculated angle θ by computing the plane (gray) spanned by the cross product of v and $\hat{k}$ (light blue) and $\hat{k}$. The positive direction is toward the vector v on the left side of the plane.

Figure 2.3: Process from the motion capture data to the angle between the difference vector v and the vertical direction $\hat{k}$

functions provided by [62], we converted this data into a set of trajectories: (x,y,z) for each joint. To further simplify the analysis, we focused on the movement of the core: more specifically, how the lower neck joint moved with respect to the pelvis (root) joint (specifically the lowerneck marker with respect to the root marker). For each dataset, we have a root (r) and lower neck (n) signal, both in $\mathbb{R}^3$. A visual representation of the vector v (green), root (yellow), and lowerneck (purple) for a single time-step are shown in Figure 2.3a superimposed onto a motion capture skeleton.

Each of the four motion capture datasets was of slightly different length, so we first resampled each r and n signal to be of uniform length T using the MATLAB `spline` function. These trajectories were not oriented in any specific way in space, so we next applied a rotation matrix to the signals to orient them in a manner that resembled our intuition about how the motion was carried out. A simple rotation by 90° about

the x-axis and flipping the signals from right to left ($r_i = r_{T-i}$ and $n_i = r_{T-i}$, $\forall i = 1, 2, ..., T$) allowed the subjects' direction of motion to be in the positive x-y direction and the vertical direction to be in the positive z direction.

Figure 2.3b shows a r (yellow) and n (purple) signal as an example. Note that the n signal is rotated counterclockwise about the r signal, so the root and neck signals are not positioned vertically above each other. We do not correct this rotation because for each dataset this offset rotation is different, and we did not wish to hand-label any of the analysis. However, this offset will be accounted for later in our analysis through a normalization process.

Next, we calculated a difference signal v which is simply $n - r$ and is the neck signal with respect to the root signal. This captures the three dimensional movements of the individual's torso during the movement and removes any overall translation effects. $v \in \mathbb{R}^3$ for one time step is shown in Figure 2.3b as the green vector, going from the yellow root signal to the purple neck signal.

The angle θ is measured with respect to the unit vector in the positive z-direction ($\hat{k}$), shown in black in Figure 2.3c). Using the relationship between the dot product and the cosine, in Equation 2.1, we find the angle θ between the vector v and the unit vector in the z-direction $\hat{k}$. When taking the inverse cosine in Equation 2.2, we ensure that the resulting angle is $0 \leq \theta \leq 90°$. The final step is the normalization of θ to $\hat{\theta}$ by subtracting the mean of θ at each time step from θ to obtain a signal centered at 0 (Equation 2.3). When performed for each dataset, this eliminates the effects of the different offset rotations and allows us to compare the oscillatory patterns between $\hat{\theta}$ signals.

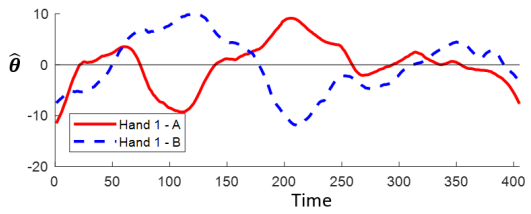
$$\|\hat{k}\|\|v\| \cos\theta = \hat{k} \cdot v \tag{2.1}$$

$$\theta = cos^{-1}\Big(\frac{\hat{k} \cdot v}{\|v\|}\Big) \tag{2.2}$$

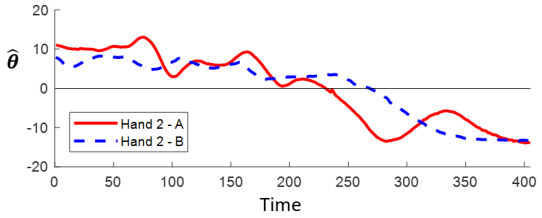
$$\hat{\theta} = \theta - \bar{\theta} \tag{2.3}$$

Because of the offset rotation explained previously, the magnitude of θ will be positive ($0 \leq \theta \leq 90°$) for all time steps. The positive direction is defined, as shown in Figure 2.3c as on one side of a plane (gray in the figure) defined as the span of $\hat{k}$ and v x $\hat{k}$ (light blue vector). The offset is different for each dataset, so the magnitude of the positive angle that represents the vertical will be different for each dataset, but characterizing the changes in the angle will show, in all cases, the oscillatory behavior of the individual's torso. Another important point is that approximating the three dimensional vector v with a single dimension will necessitate that changes in verticality in lateral and forward direction are not differentiated.

To compute the correlation between two signals, we used the Pearson correlation coefficient that takes

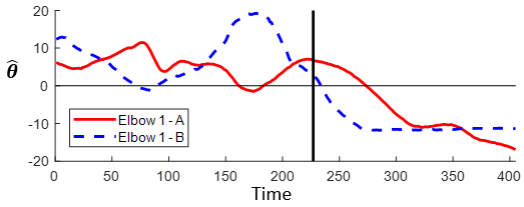


(a) Hand 1

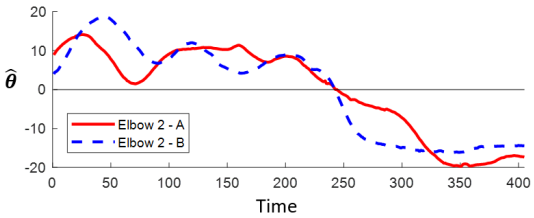


(b) Hand 2

Figure 2.4: Comparison of the verticality angle $\hat{\theta}$ (in degrees) for the Hand datasets with Subject A (red, solid) pulling and Subject B (blue, dotted) being pulled.



(a) Elbow 1, with the black line indicating the separation between high and low resistance behavior



(b) Elbow 2

Figure 2.5: Comparison of the verticality angle $\hat{\theta}$ (in degrees) for the Elbow datasets with Subject A (red, solid) pulling and Subject B (blue, dotted) being pulled.

two one dimensional signals as input and outputs the correlation between them [63]. Equation 2.4 shows the value of the correlation c for two signals, x and y, of length n.

$$c(x,y) = \frac{\sum_1^n (x-\bar{x})(y-\bar{y})}{\sqrt{\sum_1^n (x-\bar{x})^2}\sqrt{\sum_1^n (y-\bar{y})^2}} \tag{2.4}$$

2.3 Results from Verticality Analysis

For each of the four videos, we computed the $\hat{\theta}$ signals for Subject A and Subject B and compared the two signals for each video with each other. Figure 2.4 displays the Hand datasets (the two videos where the attachment point of Subject A to Subject B was Subject A's hand), and Figure 2.5 displays the Elbow datasets (where the attachment point was the elbow).

The first observation about the $\hat{\theta}$ signals for each subject in the same task (the red solid and blue dashed lines plotted together) is that there seems to be a correlation or anti-correlation between the signals. For Figure 2.4a and the first half of Figure 2.5a, the two signals seem to the anti-correlated, with oscillations in opposite directions throughout the movement. In Figure 2.4b, Figure 2.5b, and the second half of Figure 2.5a, the signals seem to be more correlated, with approximately matching shapes.

These observations match the overall structure of the comments made by our movement expert (Table

Table 2.2: Pearson correlation between $\hat{\theta}$ signals from each dataset. Also displays the resistance level and corresponding expected sign of the correlation. The Elbow 1 dataset is split into the first half (beg) and second half (end).

Video	**Resistance Level**	**Expected Sign**	**Correlation**
Hand 1	High	Negative	-0.5636
Hand 2	Low	Positive	0.8577
Elbow 1 (beginning)	High	Negative	-0.8658
Elbow 1 (end)	Low	Positive	0.702
Elbow 2	Low	Positive	0.9038

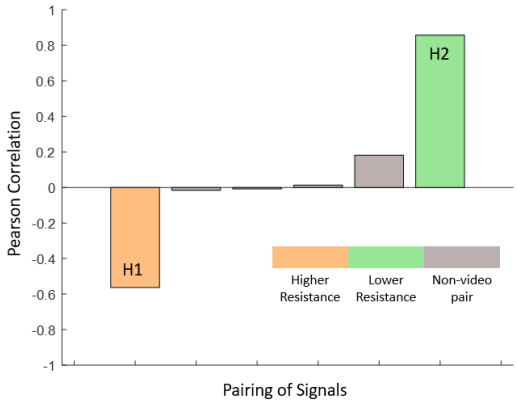
2.1). For the Hand 2 and Elbow 2 videos, there was less resistance to the pulling, manifesting in a direct correlation between the $\hat{\theta}$ signals. For the Hand 1 video, there was more resistance, manifesting in an inverse correlation between the signals. The Elbow 1 video is a special case because the movement starts out as high resistance (first half is anti-correlated), but ends as low resistance (second half is correlated).

To quantify the correlations between the signals, we computed the Pearson correlation between the signals in each video, values shown in Table 2.2. To determine the location to split the Elbow 1 signals (shown as a black line in Figure 2.5a), we found that the maximum negative correlation occurred with the first 227 points and split the signals in two parts according to that line. The table shows a high magnitude of correlation (all above 0.5) between verticality signals of two individuals performing an action together. Additionally, the sign of the correlation corresponds to the comments made by an expert about high and low resistance movements. The positive correlation values represent low resistance behaviors, and the negative correlation values represent high resistance behaviors.

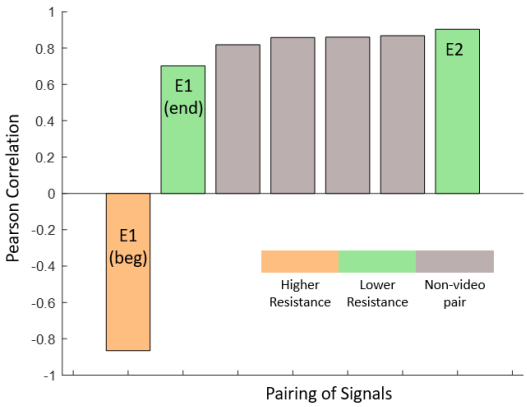
We would like to determine whether these correlation values actually indicate that two individuals are performing a task together. We will define a *video pair* as a pair of signals from the same task (i.e. Hand 2, Subject A and Hand 2, Subject B). We anticipate that the magnitude of correlation between a video pair will be higher than a *non-video pair*, which would be comparing signals extracted from two different videos. For the Hand videos, we compared all possible pairings of the 4 signals (a total of 6 pairings), which resulted in 2 video-pairs (from the two videos) and 4 non-video pairs.

Figure 2.6a displays all 6 correlation from the Hand pairings from lowest to highest with the two video pairs labeled as H1 and H2. These results match our expectations: the high and low resistance cases (video pairs) have high correlation magnitudes, and all other pairings have low magnitudes. This would indicate that a high correlation magnitude corresponded to coordination in a task, for these cases.

We performed exactly the same analysis on the Elbow videos, but with a total of 7 pairings. We compared



(a) Correlation of video and non-video pairs for Hand pairings.



(b) Correlation of video and non-video pairs for Elbow pairings.

Figure 2.6: Correlation of video and non-video pairs for Hand and Elbow signals. The bars are sorted from least to greatest with high resistance (orange), low resistance (green), and non-video pairs (gray) appropriately colored.

each of the 4 signals in pairs to each other, but then replaced the one value from the Elbow 1 video pair with two values (one from each half of the signals).

The results are shown in Figure 2.6b and do not have the same clear distinction between video and non-video pairs as in the Hand case. In fact, all non-video pairs have high correlations (above 0.8). We hypothesize that this distinction is due to the difference in point-of-contact during the interaction, and the more proximal attachment created less variety in the verticality signals within the Elbow datasets. Despite not being able to distinguish video and non-video pairs by correlation for these cases, we can still differentiate high and low resistance behavior by the sign of correlation.

2.4 Results from Resistance Analysis

From the results of the correlation calculations, there is a clear difference between the high (anti-correlated) and low (correlated) resistance cases that matches the expert observations of the movements. Another visual difference in the shapes of the signals are the varying height and oscillations between high and low resistance.

To quantify this difference, we performed a statistical analysis of each $\hat{\theta}$ signal. We first found the linear regression by the least-squares method of each signal, shown in Equations 2.5 and 2.6, where the Pearson correlation (c) of the time t and $\hat{\theta}$ are used as well as various standard deviation (SD) measures. This gave us a line about which the deviation of the signal was minimal. We then found the standard deviation of the residuals, shown in Equation 2.7.

This resulting a for each signal quantifies the amount of oscillation occurring about a line that minimizes

that oscillation. Figure 2.8 shows the linear regression for each of the 8 $\hat{\theta}$ signals separated by resistance and point of contact.

$$b = c(t, \hat{\theta}) \frac{SD(\hat{\theta})}{SD(t)} \tag{2.5}$$

$$\hat{\theta}_{est} = bt + (\bar{\hat{\theta}} - b\bar{t}) \tag{2.6}$$

$$a = SD(\hat{\theta}_{est} - \hat{\theta}) \tag{2.7}$$

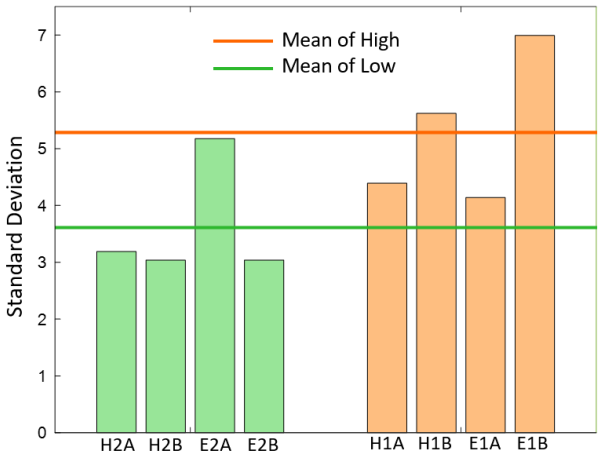
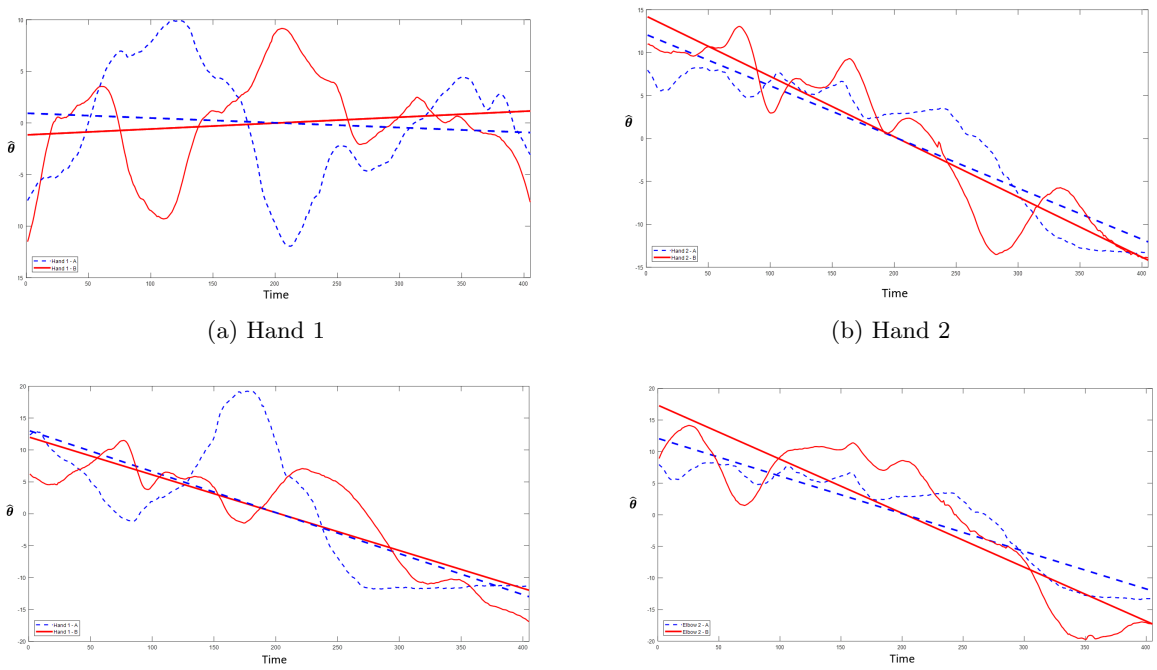


Figure 2.7: Comparison of the standard deviation of the difference of each signal from its linear resistance, separated by higher (green) and lower resistance (orange). Also pictured is the mean of the each subgroup, which shows the higher mean of the higher resistance group. The labels along the x-axis correspond to hand or elbow (H/E), video number (1/2), and subject (A/B).

To quantify the visual differences in the verticality graphs of the high and low resistance cases, we computed the standard deviation of the difference of each signal from its linear regression (from Equation 2.7). Figure 2.7 displays the standard deviation values separated by low (orange) and high (green) resistance signals. In this analysis, we have classified Elbow 1 as higher resistance for simplicity. Additionally, the mean of each group is plotted onto the figure in the appropriate color, showing that the mean standard deviation of the high resistance signals is higher than that of the low resistance signals. This agrees with our observations of the greater oscillations in the high resistance cases that quantify the differences observed by our expert about these movement pattern.



(a) Hand 1

(b) Hand 2

(c) Elbow 1

(d) Elbow 2

Figure 2.8: Linear Regression for each of the 8 $\hat{\theta}$ verticality signals with Subject A (red, solid) pulling and Subject B being pulled (blue, dashed). Classified here by point of contact (hand or elbow).

2.5 Verticality Analysis Summary

Movement of two physically connected humans is a complex activity involving several degrees of freedom, many of which motion capture does not encompass. However, we propose a model that, for our current dataset, proves to be descriptive. Our results show a correspondence between computations performed on the reduced degree-of-freedom model of the changing verticality of the subjects (a correlation metric) and features of the dyad coordination. Both whether or not two individuals were engaged in the same task with hand-to-arm contact and the quality of "resistance" in the partners were considered. Our analysis corroborates the naïve observations a movement expert. Happily, our analysis does not depend on the type of task, so the verticality metric can be easily tested on a variety of other tasks and contexts to determine other conclusions that can be drawn from this measure.

Chapter 3

Low-DOF Artificial Systems and Modeling for Human Imitation

In this chapter, we will first explain our process of modeling motion capture data with four degrees of freedom (Section 3.1). We then discuss the process of artificially generating the orientation of these simulated robots (Section 3.2. We will conclude this chapter by explaining the explain the correspondences between this model and the movement of two simulated robots, the Rollbot and Broombot, in Section 3.3. The code for these simulations is written using Python 3 and the most updated version is hosted at `https://github.com/roshk99/Robot-Menagerie`. The Readme file for this repository is included Appendix A.

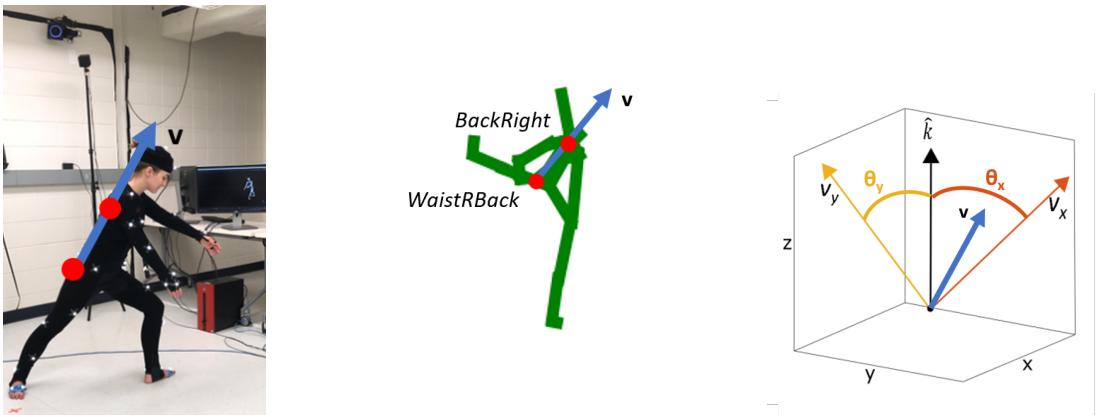
3.1 Computing a 4-DOF Model of Human Motion

We first recorded human motion data using an OptiTrack system and the Baseline markerset containing 37 markers (Figure 3.1a). We exported the data from the Motive software in the form of Cartesian coordinates containing the X-Y-Z translation of each marker for each step in time. The frame rate of capture was 120 Hz, or a time step of 0.0083 seconds. Before performing further computations in the data, we manufactured the data at any points where the motion capture system lost the position of any of the markers by linearly interpolating between successfully recorded values.

The first two components of this model are the translation coordinates (x, y), representing the overall human translation in the x-y plane. We obtained these values by using the x and y coordinates of the marker *BackTop* (or m_0) at every time step (Equation 3.1).

$$\begin{aligned} x &= m_{0,x} \\ y &= m_{0,y} \end{aligned} \tag{3.1}$$

We next chose a pair of markers from the dataset to create a vector corresponding to some aspect of the skeletal motion. The vector, v, between any marker pair, (m_1, m_2), will have two planar projections, providing two additional, rotational degrees of freedom. Thus, we form a human motion model with two translational degrees of freedom (common to any choice of marker pairs) and two rotational degrees of



(a) A vector v (blue) between two motion capture markers on a human wearing a motion capture suit.

(b) A vector v (blue) between two motion capture markers, m_1 (top red dot) and m_2 (bottom red dot).

(c) Illustration of the projections v_x (orange) and v_y (yellow) and the angles θ_x and θ_y for the same vector v (blue) at a single time step.

Figure 3.1: Human motion capture data to a four degree-of-freedom model, with the two orientation components shown here

freedom, for which we are interested in measuring human preference for across a series of data samples.

For example, Figure 3.1 shows the details of one possible mapping. This mapping uses the *WaistRBack* and *BackRight* markers, which is a representation of overall spinal leaning that we call verticality; thus in this case $m_1 = WaistRBack$ and $m_2 = BackRight$. The vector v, shown in blue in Figure 3.1b, is computed by subtracting the (x,y,z) coordinates of the initial point (m_1) from the coordinates of the end point (m_2) to obtain a vector that represents the vertical leaning of the motion capture skeleton at each time step. This vector v is then normalized at each time step to $\bar{v} = v/||v||$ to characterize the direction of the vector v (Equation 3.2).

$$\begin{aligned} v &= m_2 - m_1 \\ \bar{v} &= v/||v|| \end{aligned} \tag{3.2}$$

In the following equations, $\hat{i}, \hat{j}$, and $\hat{k}$ are used to represent the unit vectors in the x, y, and z directions, respectively. From the 3 dimensional signal $\bar{v}$, we will obtain angles that characterize the rotation about the x-axis and y-axis (Figure 3.1c). The axes are aligned such that the x-y plane is the ground. To calculate the angles in these two directions, we first calculate the projection of vector $\bar{v}$ at each time step on the y-z plane (v_x) and the x-z plane (v_y) (Equation 3.3). Equation 3.4 shows the calculation of θ_x and θ_y, which are illustrated in Figure 3.1c as the angles between the projected vectors and the z-axis. The signs of these angles are adjusted to ensure they match the right-hand-rule.

$$
\begin{aligned}
v_x &= \hat{i} \times (\bar{v} \times \hat{i}) \\
v_y &= \hat{j} \times (\bar{v} \times \hat{j})
\end{aligned}
\tag{3.3}
$$

$$
\begin{aligned}
\theta_x &= sgn(v_x \cdot (\hat{k} \cdot v_x))|\cos(\hat{k} \cdot v_x)| \\
\theta_y &= sgn(v_y \cdot (\hat{k} \cdot v_y))|\cos(\hat{k} \cdot v_y)|
\end{aligned}
\tag{3.4}
$$

Combining the results from the translation x, y from Equation 3.1 and the rotation θ_x, θ_y from Equation 3.4, we have a total of 4 degrees of freedom for our model of human motion.

3.2 Artificially Generated Motion

We developed a method to artificially generate the orientation of the Rollbot and Broombot to compare to the motion-capture-based method developed above. We only modified the rotation components of the robot motion (i.e. $\theta_x, \theta_y, \theta_z$) and left the translation components as mirroring the horizontal movement of the human.

We generated a new value for each of the angles at each time step by first computing random amplitudes A between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$. We then used Equation 3.5 to calculate the angle θ_i. This calculation can be performed for the Rollbot for $i = z$ and for the Broombot for $i = x$ and $i = y$. The constant $\omega = \frac{1}{5000}$ was found by hand-tuning and was chosen so the resulting angles would approximately match the speed of movement of the motion capture data.

$$
\theta_i = A_i \sin(\omega t) \tag{3.5}
$$

3.3 Commanding the Simulated Robots

Once we have created a model for the motion capture data, we designed the correspondence between this model and the motion of the simulated robots. We have created two low degree of freedom simulated robots, with three and four degrees of freedom, pictured in Figure 3.2. The dual colors of each robot make the rotations more apparent when viewed in the simulation. The first robot, Rollbot, can translate in the x and y directions in the horizontal plane and has one rotation θ_z about its central axis. The second, Broombot, can also translate in the x and y directions in the x-y plane and has two rotations: one about its x-axis and

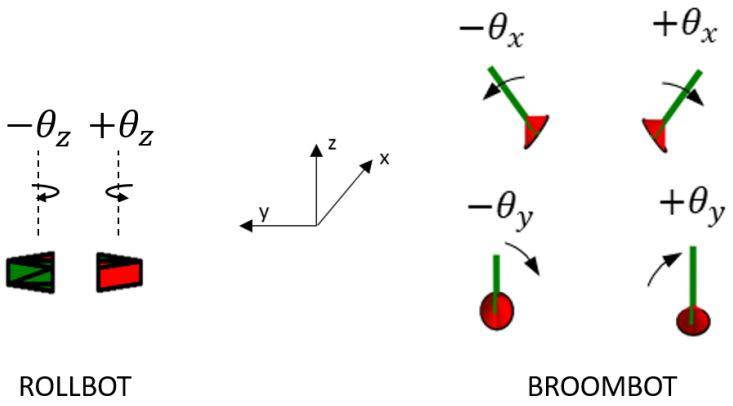


Figure 3.2: The Rollbot (**Left**) has one articulated degree of freedom, θ_z, that is a rotation about the z-axis. The Broombot (**Right**) has two articulated degrees of freedom, θ_x about the x-axis and θ_y about the y-axis. With this coordinate frame orientation, this is accomplished by tilting left to right and front to back. Both robots can translate horizontally in the x-y plane.

Table 3.1: Five different mappings with different starting and ending markers for the vector v used to model the human motion. The sixth mapping is artificially generated motion, not based on human motion.

	Marker 1 (m_1)	**Marker 2 (m_2)**	**Related Body Part**
Mapping 1	*ShoulderLBack*	*WristLOuter*	Arm Left
Mapping 2	*ShoulderRBack*	*WristROuter*	Arm Right
Mapping 3	*WaistLBack*	*AnkleLOuter*	Leg Left
Mapping 4	*WaistRBack*	*AnkleROuter*	Leg Right
Mapping 5	*WaistRBack*	*BackRight*	Spine
Mapping 6	Artificially Generated - See Section 3.2		

one about its y-axis.

The Broombot's four degrees of freedom correspond naturally to our four DOF model. We will command the translation of the Broombot with translation components of the model and the rotation of the Broombot with rotation components of the model. The only modification will be to offset the translation with a constant at every time step so the Broombot and motion capture skeleton plotted simultaneously will translate synchronously with starting from an offset position.

The Rollbot's three degrees of freedom require a slight modification to use our model. The translation will correspond directly to the translation components of our model (with an offset as in the Broombot), but the two rotation components in the model must correspond to a single rotation of the Rollbot. We add

these two rotations together (i.e. $\theta_z = \theta_x + \theta_y$) to achieve this.

We can use a variety of pairs of markers to command the Broombot's motion, but we chose a set of five marker pairs that correspond to to various connected components of the body. These five mappings from human motion to robot motion are shown in Table 3.1 with the starting (m_1) and ending (m_2) points of the vector v listed.

Chapter 4

Imitation of Human Motion using Low-DOF Simulated Robots

This chapter describes the experimental setup and results of two studies investigating how users perceived the imitation of human motion by low-DOF simulated robots. An initial study, presented in Section 4.1 explores the imitation success of the Rollbot and Broombot (introduced in Section 3.3) using verticality, or the leaning of the spine, of motion capture data. Section 4.2 details the results of this survey. We next created an experiment exploring how user imitation preferences for Broombot motion change as the motion capture sample and mapping to human motion is changed (Section 4.3). A discussion of these results is presented in Section 4.4.

4.1 Simulated Robots Imitating Human Motion with Verticality

Experimental Design

We recorded five minutes of motion capture data, recorded via the OptiTrack motion capture system. The subject [1] moving in the motion capture suit was a researcher in our lab who is trained in classical ballet, and this researcher moved freely to create a variety of movement for the five recorded minutes.. We then chose three different samples of the recorded movement to use for robot motion generation. For each of these three motion capture samples, we generated three different cases of robot motion, enumerated below. The verticality mapping used in this study is *Mapping 5* from Table 3.1, and the artificially generated mapping is *Mapping 6* from Table 3.1). A still image from an animation is shown in Figure 4.1.

We had three different case formats as shown below. We generated 3 videos for Cases 1 and 2 and 2 videos for Case 3, but also created 8 mirrored videos that switched the positioning of the movers in each video. Each user was presented with a total of 8 questions in a random order.

[1] We would like to thank Erin Berl for agreeing to be the subject for these studies

Figure 4.1: Still image from an example animation for the study with three movers (two simulated robots and one human motion capture skeleton). The robots were driven by either the verticality mapping (*Mapping 5* from Table 3.1) or the artificially generated mapping (*Mapping 6* from Table 3.1), and users were asked which robot better imitated the human movement.

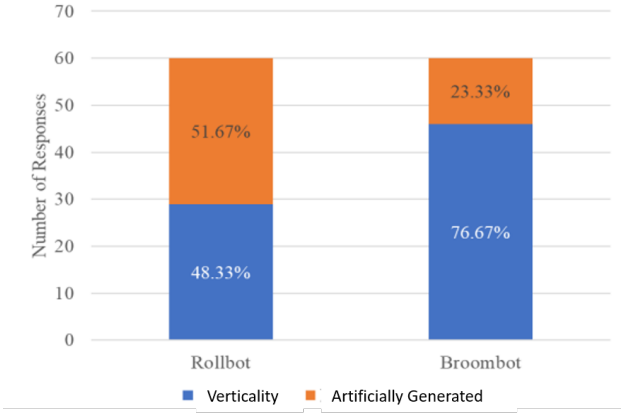
- Case 1: Rollbot Verticality vs. Artificially Generated
- Case 2: Brooombot Verticality vs. Artificially Generated
- Case 3: Brooombot Verticality vs. Rollbot Verticality

For each of the 8 questions, the user was randomly shown either the original or mirrored video and asked two questions. They were first asked "Which robot (left or right) imitates the human skeleton (center) best?" and could pick either Left or Right. They were then prompted to explain in a short answer form why they made their choice for the first question.

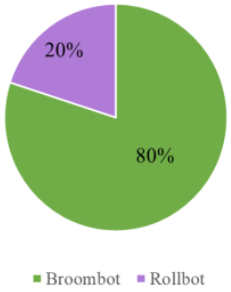
Through these questions we aimed to test two hypotheses. **Hypothesis 1** was that users prefer verticality over artificially generated motion for the Rollbot and Broombot (corresponds to Cases 1 and 2). **Hypothesis 2** states that when both robots use verticality, users prefer the Broombot over the Rollbot for imitation (corresponds to Case 3).

4.2 Simulated Robots Imitating Human Motion with Verticality Results

This study was developed using SurveyMonkey and administered using Amazon Mechanical Turk (MTurk) to 20 online participants. We first compared the number of participants identifying the verticality-based motion as imitating the human skeleton movement better to the number preferring the artificially generated motion (Figure 4.2a). In the case of Rollbot motion (Case 1), the users had no preference between the verticality and artificially generated robot motion. Some excerpts from their justifications include: "They are almost identical" and "I could not tell any difference in the way the robot moved." However, in the



(a) Number of responses identifying the verticality-based robot motion (blue) and artificially generated robot motion (orange) as imitating the human skeleton movement best for the *Rollbot* and *Broombot* movements.



(b) Percentage of participants identifying the *Rollbot* verticality-based motion (purple) and the *Broombot* verticality-based motion (green) as imitating the human skeleton best (total of 40 responses).

Figure 4.2: Results from study comparing verticality and artificially generated motion on the Rollbot and Broombot

case of *Broombot* motion (Case 2), the users preferred the verticality motion 77% of the time, with one user commenting: "This broom seems to be moving exactly like the human skeleton."

We then compared the verticality based motion between the Rollbot and Broombot in Case 3 (Figure 4.2b). The users preferred the Broombot 78% of the time, with users commenting: "overall movement is more active and copies movements better" and "it can emulate the leaning of the upper torso."

We then performed one-sample two-sided t-tests on all three datasets (Table 4.1), comparing the resulting percentages with a "random" choice of 50%. The degrees of freedom were $60 - 1 = 59$ for Cases 1 and 2 and $40 - 1 = 39$ for Case 3.

For Case 1, the two-sided t-test was not significant ($p > 0.01$), so the probability of choosing verticality over artificially generated motion cannot be distinguished from 0.5 for the Rollbot. For Case 2, both the two-sided and one-sided t-tests were significant at the 0.01 level, showing that users preferred the verticality motion over the artificially generated motion for the Broombot. This only confirms **Hypothesis 1** for the Broombot and not for the Rollbot, since users preferred verticality only for the Broombot motion.

For Case 3, both the two-sided and one-sided t-tests were also significant, illustrating the users' preference of the Broombot over the Rollbot when both were using verticality. This confirms **Hypothesis 2** since the users preferred the higher DOF Broombot over the Rollbot when both were moving using verticality.

In this study, we generated motion profiles of virtual characters imitating human movement by utilizing

Table 4.1: P-Values for One Sample T-Tests on User Study Results ($p = 0.01$ significance)

	Two-Sided T-test	**One-Sided T-Test**
Verticality Rollbot	0.799	0.601
Verticality Broombot	9.622e−6	4.811e−6
Verticality Broombot vs. Rollbot	3.379e−5	1.689e−5

a verticality metric computed from motion capture. This verticality metric, drawing inspiration from motion capture analysis of dyadic motion (Chapter 2), was applied to the creation of 3 and 4 degree of freedom virtual characters (Chapter 3). Users evaluated the motion of these low degree of freedom characters with respect to their imitation capabilities of human movement, comparing verticality and an artificially generated mapping. Statistical analyses on the user responses illustrated a preference for the verticality-based motion and the higher degree of freedom virtual character Broombot.

4.3 Human Perception and Preference of Low Degree-of-Freedom Simulated Robot Motion Driven by Various Human Capture Mappings

The previous sections detailed a preference for verticality-based motion and the Broombot's motion for specific examples of motion capture data. Based on this initial study, we developed a new study to investigate how users' preferences change as the motion capture samples are varied and mappings other than verticality are used for imitation.

In this section, we discuss the choice of motion capture data clips, or *samples*, (Section 4.3.1) and the choice of the *mappings* to command the Broombot (Section 4.3.2) for the user study. We then explain the composition of the stimuli by combining the motion capture samples and the mappings to command the Broombot. We explain the development of the user study using these stimuli (Section 4.3.3). The individual components of the survey include the pre-survey training (Section 4.3.4), questions for each stimuli (Section 4.3.5), and collection of demographic information (Section 4.3.6).

4.3.1 Choosing Motion Capture Samples

We recorded five minutes of motion capture data, recorded via the OptiTrack motion capture system (the same recording as the initial study in Section 4.1). The subject moving in the motion capture suit was a researcher in our lab who is trained in classical ballet, and this researcher moved freely to create a variety of movement for the five recorded minutes. We observed this recording and chose three samples that focused

on movement from on a specific section of the body, which are enumerated below.

- **Sample A**: 25 second clip of motion capture observed by the researchers to have high deflections in arm movements
- **Sample B**: 25 second clip of motion capture observed by the researchers to have high deflections in leg movements
- **Sample C**: 25 second clip of motion capture observed by the researchers to have high deflections in spinal movements

4.3.2 Choosing Mappings to Command Robot Motion

We used six different mappings from human motion to the Broombot's motion (Table 3.1). The first five mappings use five vectors from the motion capture skeleton. The designations of left and right throughout this paper are in relation to the mover in the video, not the viewer of the video. An illustration of *Mappings 2, 4,* and *5* on the motion capture skeleton and the corresponding position and orientation of the Broombot are shown in Figure 4.3. The sixth mapping has an artificially generated orientation for the Broombot, via the process from Section 3.2.

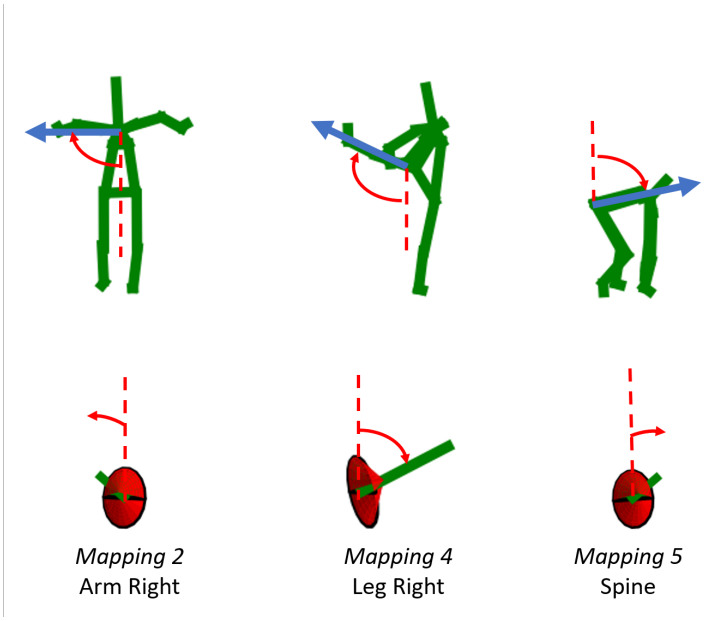


Figure 4.3: Three possible mappings from human to robot motion. The top row of the figure displays the vector (blue) on the motion capture skeleton and the axis with respect to which the two angles are calculated. Those two angles (rotation about the the x and y axes) correspond to the orientation of the Broombot (second row).

4.3.3 Developing the User Study

For each stimuli, we generated a video that included a motion capture skeleton in the center with one Broombot on either side of it. A screenshot of a video is shown in Figure 4.4. With the six mappings, this generated a total of $\frac{6!}{(6-2)!} = 30$ pairs to compare. We used those 30 combinations for each of the three motion capture samples, resulting in a total of 90 videos.

We split these 90 videos into four groups, two groups with 23 videos and two groups with 22 videos. We selected the videos in each group to ensure that each group saw videos from all three samples and all six mappings. We also ordered the videos within each group so the respondents did not see the same mapping or sample multiple times in a row.

We used the 90 videos to run a user study via a Qualtrics survey administered through Amazon Mechanical Turk (MTurk). The goal of this user study was to determine the user preferences of the robot imitation success of a variety of mappings and motion capture samples. We developed a series of hypotheses, listed below, that will be addressed through the results of the study.

- **Hypothesis 1**: Users will prefer mappings based on motion capture (*Mappings 1-5*) over artificially generated motion (*Mapping 6*).
- **Hypothesis 2**: Users will be more likely to choose mappings from arms (*Mappings 1-2*) for **Sample A**, legs (*Mappings 3-4*) for **Sample B**, and spine (*Mapping 5*) for **Sample C**.
- **Hypothesis 3**: Spinal movements (*Mapping 5*) will be chosen more often than other mappings. This hypothesis is based on the preference users showed for verticality in a previous study and the correlation verticality had with dyadic movement.

The respondents were first presented with a consent form and then proceeded to pre-survey training (Section 4.3.4) that prepared them to compare the imitation success by two Broombots of a motion capture skeleton. They next were randomly sorted into one of four groups (from Section 4.3.3) and presented with one question for each of those videos (Section 4.3.5). They completed the survey with demographic information (Section 4.3.6) and a random number they entered into the MTurk website for completion. Upon confirmation of completion, each respondent was paid $1, a rate determined from an estimated 30 minutes for completion.

4.3.4 Pre-Survey Training

We included a training screen before each respondent began answering questions about the stimuli. The purpose of this training was to (1) show imitation of human motion as different from replicating exactly, (2)

demonstrate examples of non-humanoid characters successfully imitating human motion, and (3) familiarize respondents with the appearance and movement patterns of the motion capture skeletons and simulated robots Broombot.

We first displayed definitions and examples of imitation and copying so respondents were aware that imitation does not require movers to look or move exactly the same. These definitions are reproduced here:

- Imitation: The two movers are moving similarly, and do not necessarily look the same. A dog can imitate a human, even though they look very different. For imitation the movements must resemble each other.
- Copying: The two movers look exactly the same and move exactly the same for the entire movement.

We next included an example of three dancers from a YouTube video performing the same choreography to showcase that the dancers are moving similarly and are imitating each other; however, they do not look the same and are not moving exactly alike.

We included a YouTube video of a human dancing with a cartoon mouse character to show that imitation does not require the appearance of the two movers to be the same. This again illustrated the differences between copying and imitating and allowed the participants to see that our robots could look very dissimilar to humans and still imitate their movement.

The third question was worded in the exact way each video question in the survey was phrased. Participants were shown a video of three motion capture skeletons, in which the left mover was imitating the center mover but the right mover was moving very differently. They were required to answer this question correctly (i.e. choose Left) to proceed to the survey.

The final section of the training process introduced the respondents to the Broombot simulated robot and ways it can both translate and tilt. At the very end of training, the respondents confirmed that they were required to watch the videos completely and write an explanation for each of their answers.

After completing this training, the respondents were more familiar with the task they needed to complete and understood our definitions of imitation. We sought to avoid the scenario where respondents merely discarded the imitation capabilities of the simulated robots due to their lack of humanoid appearance and wanted respondents to evaluate the robots based on their movements.

4.3.5 Questions for Each Video

For each video, we asked up to three questions. A screenshot with a sample of the screen each respondent saw for a video question is shown in Figure 4.4. The first question after watching a video asked the respondents:

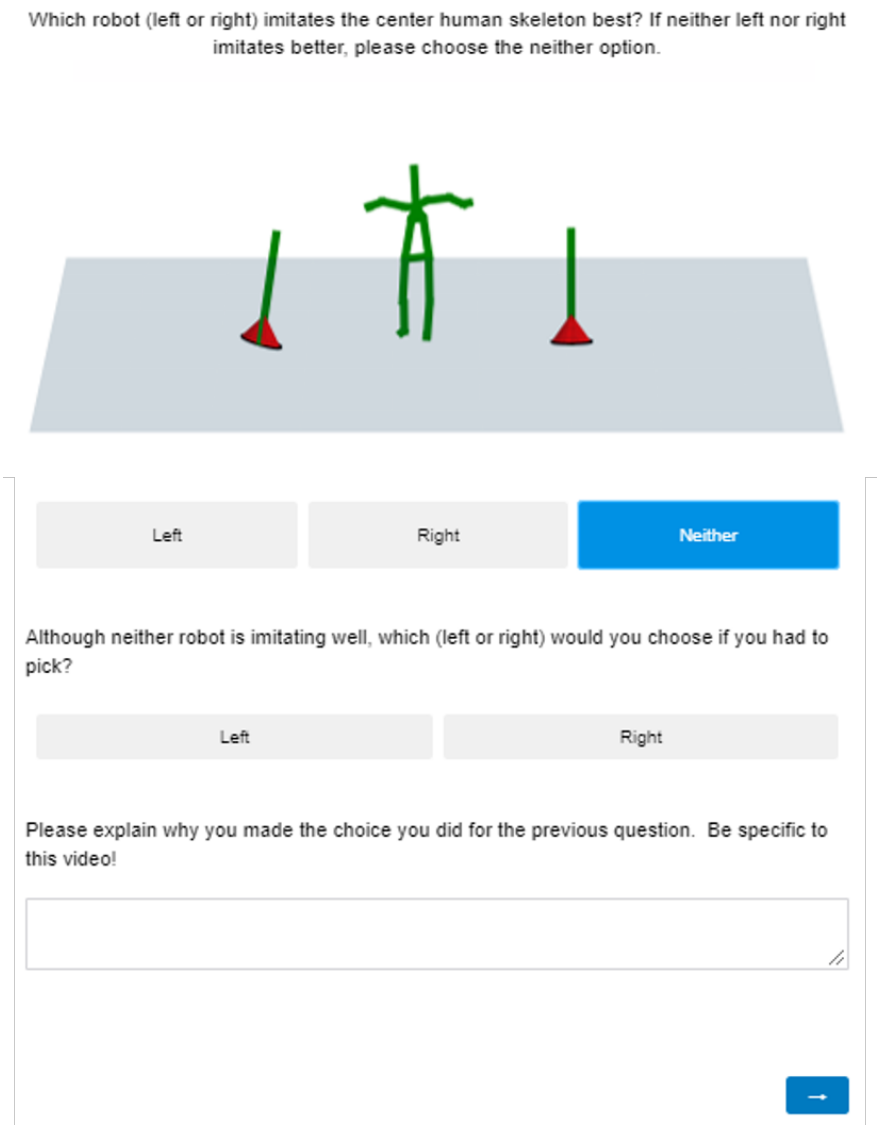


Figure 4.4: An example of a question from the user study that asks respondents to watch a video containing two simulated robots imitating a motion capture skeleton and determine which robot is imitating the skeleton's movement better. The second question forcing a response of Left or Right only appeared for respondents who chose Neither for the first question. Respondents were also required to provide an explanation of their response via short answer question with a minimum length.

"Which robot (left or right) imitates the center human skeleton best? If neither left nor right imitates better, please choose the neither option." The respondents were provided with three choices for this question: Left, Right, and Neither. This is similar to the question asked in the initial study (Section 4.1), but includes a Neither option instead of forcing respondents to choose between Left and Right.

If they chose Neither as the answer to this question, a second question appeared and asked "Although neither robot is imitating well, which (left or right) would you choose if you had to pick?". This allowed us

to collect responses without a forced response (first question) as well as obtain data with a forced response (second question). The results discussed below will include data both with and without the forced response.

The final question asks for a short explanation of the respondent's choice for the previous questions by asking "Please explain why you made the choice you did for the previous question. Be specific to this video!". We required their responses to achieve a minimum length, communicated by users through an error message for answers that were too short. The goal of this was to encourage respondents to elaborate upon their explanations instead of just providing 1-2 word responses.

4.3.6 Demographic Information

All demographic information collected at the end of the survey was optional. Among the questions asked were age range, gender, and education level. We also included a table in which respondents could input their experience level (if any) in several movement based activities out of three categories: None, 0-2 years, 2-4 years, and More than 4 years. The activities displayed were running, triathlon, dance class or practice, martial arts, sports, pilates/group fitness class, swimming, walking/hiking, weight lifting, and other. The demographic information aggregated over all respondents is shown in Appendix B. Most respondents were within the age range of 25-34, spoke English as a native language, and had a bachelor's degree in college. Respondents of various movement activity experience levels were represented.

4.4 Human Preference Analysis of Imitation of Low-DOF Simulated Robots Results and Discussion

In this section, we will discuss the results from the user study. We inserted questions during the survey such as simple addition and multiplication problems to ensure respondents were keeping engaged in the survey and not just answering randomly. We eliminated the responses from four users who did not answer these questions correctly, resulting in a total of 196 respondents. We will discuss the results by motion capture sample in Section 4.4.1 and the distribution of preferences for respondents in Section 4.4.2. We will perform an analysis of the three motion capture samples from the study in Section 4.4.3 and conclude with a summary and discussion of all the results in 4.4.4.

4.4.1 Results by Motion Capture Sample

To compare the preferences for different mappings commanding the Broombot's motion, we analyzed the results from our user study by motion capture sample. After discussing our data analysis process, we will

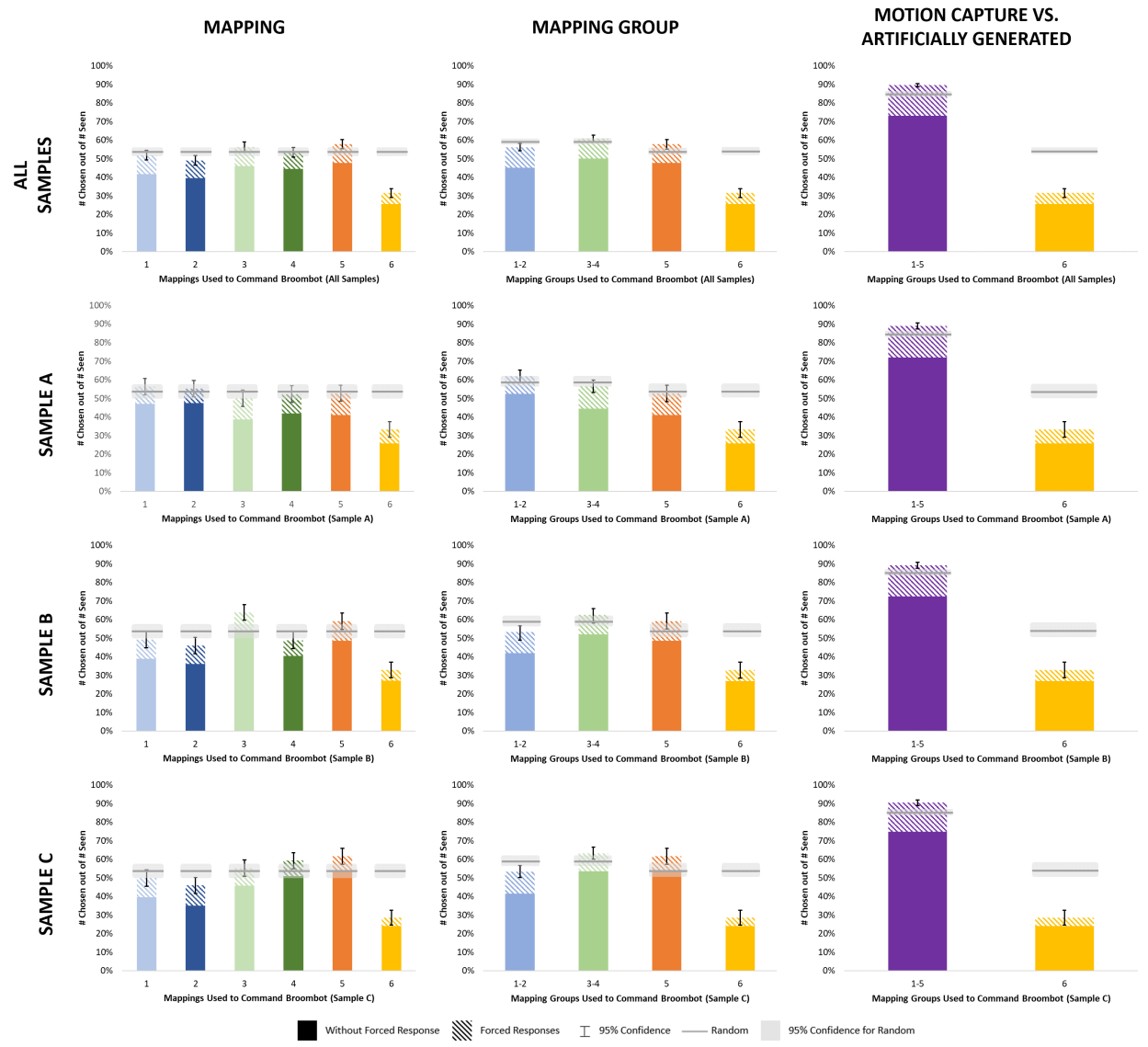


Figure 4.5: User study results separated by motion capture sample for each mapping (left column), for each mapping group (middle column), and for motion capture vs. artificially generated (right column). The gray line indicates the result if all respondents chose Left and Right randomly for each question, and the 95% confidence interval for both the results and the random choice are indicated.

review the results for each motion capture sample by individual mapping, mapping group, and motion capture vs. artificially generated.

We computed a percentage that represents the frequency a mapping (one of six) was chosen out of the number of times it was seen. Since each mapping was not seen an equal number of times, having the number of times seen in the denominator allows us to compare the percentages. The naming scheme for motion capture samples and mappings to robot motion was discussed in Section 4.3.1 and 4.3.2.

We also computed a 95% confidence interval for the resulting percentages using Equation 4.1, where $\hat{p}$ is our calculated percentage and 1.96 is the constant multiplier for at 95% confidence level. The value of the denominator, N, is the number of times a mapping was seen over all participants.

$$\hat{p} \pm 1.96 \frac{\hat{p}(1-\hat{p})}{N} \tag{4.1}$$

Figure 4.5 (left column) shows the results separated by motion capture sample and aggregated over all samples. The fully shaded portion of each bar indicates the responses without the forced response, and the striped portion is the added results with the forced responses. The gray line marks the level for each mapping if all participants chose Left and Right randomly for the entire survey, and the corresponding 95% error region is shaded gray. The 95% confidence interval was calculated for each bar, with the forced response for clarity, and is shown by an error interval.

For each sample and overall, each individual mapping was in the error range of random choice except *Mapping 6*, which respondents chose infrequently. Additionally, the responses for the left and right sides of the body (*Mappings 1-2* and *Mappings 3-4*) were approximately equal, so respondents did not prefer a mapping from the left or right.

Grouping the responses into four mapping groups (middle column), we will combine the responses into *Mappings 1-2*, *Mappings 3-4*, *Mapping 5*, and *Mapping 6*. Since the number of responses for the *Mappings 1-2* and *Mappings 3-4* groups is greater than the number of responses for *Mapping 5* and *Mapping 6*, the denominator of our calculation, number of times the mapping group was seen, will adjust accordingly.

Figure 4.5 (middle column) contains the results for each mapping group. The line indicating random choice is not at the same level for each group due to the increased opportunity to choose *Mappings 1-2* or *Mappings 3-4* when randomly choosing Left or Right for each video. For **Sample A**, *Mappings 1-2* were chosen at a frequency greater than random. For **Samples B and C**, *Mappings 3-4* and *Mapping 5* were chosen at a frequency greater than random, with *Mappings 3-4* preferred slightly more than *Mapping 5* for both samples. Overall, *Mappings 3-4* and *Mapping 5* were chosen more frequently than random, and *Mapping 6* was chosen much less frequently than random.

Grouping all motion capture mappings (*Mappings 1-5*) into one group, we can see how our motion capture mappings performed compared to artificially generated motion in Figure 4.5 (right column). For each sample individually and all samples combined, *Mappings 1-5* were chosen more frequently than random and *Mapping 6* was chosen much less frequently than random.

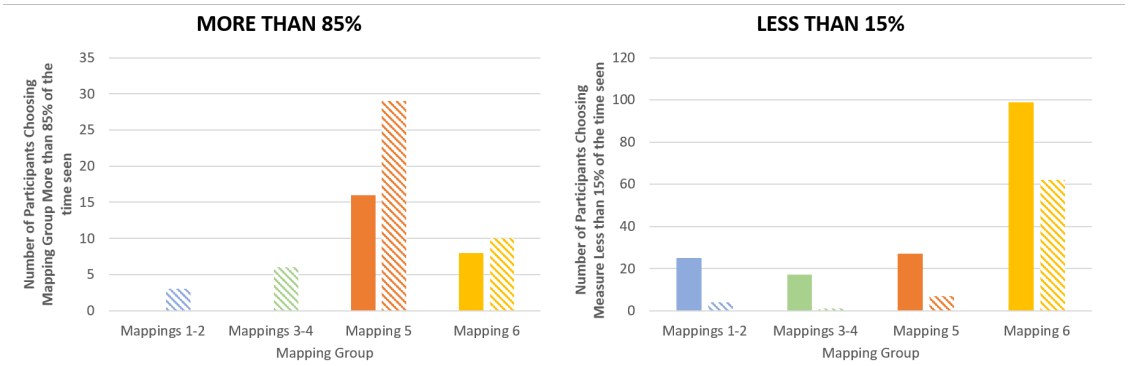


Figure 4.6: The number of respondents choosing each mapping group (*Mappings 1-2*, *Mappings 3-4*, *Mapping 5*, and *Mapping 6*) more than 85% of the times seen (left) and less than 15% of the times seen (right). The values without the forced response are fully shaded, and the values without the forced response are partially shaded.

4.4.2 Respondent Preferences Distribution

In this section, we will discuss the distribution of respondent choices for our survey separated by mapping groups (*Mappings 1-2*, *Mappings 3-4*, *Mapping 5*, and *Mapping 6*). We first counted the number of each mapping group seen and chosen by each respondent. The ratio of these values results in a percentage of times each respondent chose a mapping group given the number of times seen.

The Pearson Coefficient of Skewness was computed for the distributions of each mapping group. A number closer to 0 indicates a less skewed distribution. The groups based on motion capture (*Mappings 1-2*, *Mappings 3-4*, *Mapping 5*) had low magnitudes of skewness without forced response, with values of 0.65, -0.30, and -0.20 respectively. In contrast, the group based on artificially generated motion (*Mapping 6*) had a high skewness value of 1.33, indicating that many respondents chose artificially generated motion very infrequently. From these distributions, we plotted the number of respondents choosing a given mapping group more than 85% of the times seen and less than 15% of the times seen in Figure 4.6, to see if any mapping was chosen very frequently or infrequently.

There were many respondents who chose artificially generated motion (*Mapping 6*) very infrequently (less than 15% of the time). The remaining mappings have a low number of respondents choosing them infrequently. This result will tallies with data presented in Section 4.4.1 that shows respondents chose motion capture-based motion more frequently than artificially generated motion. Additionally, there were more respondents choosing spinal movements (*Mapping 5*) very frequently (more than 85% of the time) than any other mapping. Though this does not indicate that the mapping based on spinal movements was chosen more overall, it does indicate a high "loyalty" to verticality by some respondents (16 without forced response and 29 with forced response).

Loyalty and Engagement in Verticality Respondents

To compare the verticality respondents, or those individuals that chose *Mapping 5* very frequently, to the entire respondent pool, we developed quantitative values for the loyalty and engagement of participants within a group. We computed these values for the verticality respondent group, the respondents not in the verticality respondents group, and the total respondent pool.

Table 4.2: Loyalty values calculated for verticality respondents, not verticality respondents, and all respondents. Loyalty includes percentages for each mapping group.

	LOYALTY			
	Mappings 1-2	**Mappings 3-4**	**Mapping 5**	**Mapping 6**
Verticality Respondents	0.0%	6.2%	91.0%	0.0%
Not Verticality Respondents	18.7%	20.8%	16.3%	16.5%
All Respondents	15.9%	18.6%	27.4%	14.1%

To analyze loyalty, we first found the mappings that each participant chose the most often (using the values with forced response for this analysis). This differs from 85%,15% analysis performed in the previous section as it determines the mapping that each individual was most loyal to. For each subgroup of respondents, the loyalty measure was calculated using Equation 4.2, and the results are tabulated in Table 4.2. This calculation confirms that the verticality respondent group contains high loyalty to spinal movement mapping (91%), and the other groups do not show a particular loyalty to one mapping (highest value 27.4%). This methodology allows us to segment our participants into groups and quantify the loyalty of the members of that group to specific mappings.

$$\text{Loyalty(Group, Mapping)} = \frac{\sum \text{Mapping Preference \% for Loyal}}{\text{Number in Group}} \tag{4.2}$$

We also quantified engagement of the respondents by looking at the short answer responses. We first determined whether each respondent wrote 'relevant' responses (percentages in Table 4.3). The users of Amazon MTurk are likely to simple copy and paste a random text into required short answer questions, and determining 'relevant' responses allow us to screen out those participants' short answer responses. We set two criteria that respondents had to meet in order to ensure that they provided relevant responses to the short answer questions:

Table 4.3: Engagement values calculated for verticality respondents, not verticality respondents, and all respondents. Engagement is the multiplication of the percentage of respondents writing relevant responses and the average characters written for relevant responses.

	ENGAGEMENT		
	% Relevant	**Characters for Relevant**	**Engagement**
Verticality Respondents	57.1%	93.3	53.3
Not Verticality Respondents	34.4%	75.0	25.8
All Respondents	79.0%	79.0	29.0

- Number of responses containing at least one of the words (left, center, middle, right, human, robot) for the respondent was greater than **15** ($KW > 15$)
- Number of unique words across all responses of a respondent was greater than **20** ($UW > 20$)

For each respondent in the group, we can determine if they provided relevant responses ($KW > 15$ and $UW > 20$), and we can then calculate a percentages of respondents in each group writing relevant responses. If all respondents in a group wrote relevant responses, this value would be 100%, which will be inputted into the engagement formula (Equation 4.3).

We can additionally count the average number of characters written for respondents who wrote relevant responses. Combining these two numbers allows us to determine an engagement score using Equation 4.3 for each group, with results shown in Table 4.3. We can see that the verticality respondents were on average more 'engaged' with a score of 53.3 compared to the 29.0 over all respondents. This engagement value allows us to determine how involved respondents from a group were with our survey.

$$\text{Engagement(Group)} = \%\ \text{Relevant} * \text{Average Characters for Relevant} \tag{4.3}$$

We also investigated any other trends in the demographic information for these verticality respondents, to see if any other factors could correlate with the higher loyalty and engagement. The results discussed here are those with a difference of greater than 10% in the responses.

The verticality respondents were 10% more female than the overall sample. They had more experience in some activities (sports, swimming, and walking/hiking) and less in others (martial arts, yoga, and running), but when aggregating all the activities, differences were not greater than 10%. Other demographic responses, such as language and age, did not show significant differences in results between verticality respondents and

the entire pool of respondents.

Overall, verticality respondents, who chose *Mapping 5* more than 85% of the time, were loyal to the mapping based on spinal movements, were more likely to be female, and were more engaged in the short answer component of our survey. Further demographic analysis of these respondents may lead to greater understanding of the relationship between preferring verticality and a greater attentiveness during the survey.

4.4.3 Analysis of Motion Capture Data

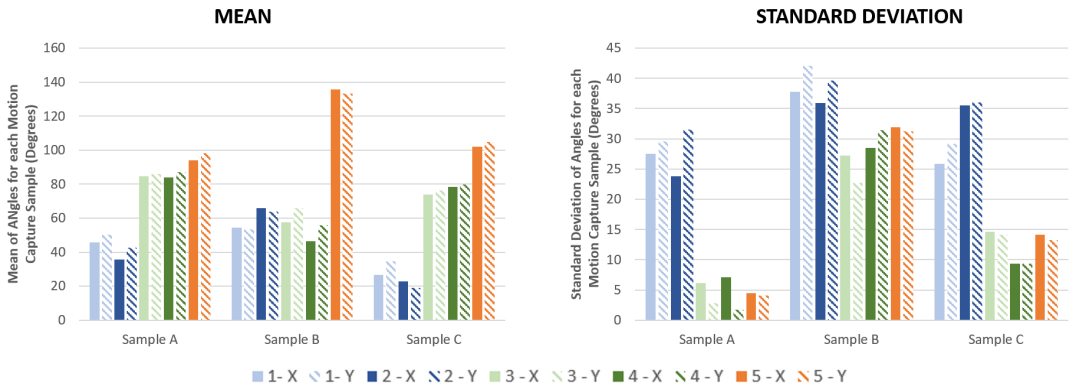


Figure 4.7: The mean (left) and standard deviation (right) of the five motion capture mappings, separated by motion capture sample. The values for the angles in the x and y direction are displayed with bars representing θ_x fully shaded and bars representing θ_y striped.

We also analyzed the angles (θ_x, θ_y) for all the five motion capture based mappings and for each motion capture sample to determine if statistical methods could correlate with the respondents' preferences in the user study. We computed the mean and standard deviation for θ_x and θ_y, shown in Figure 4.7.

Although the mean of the spine mapping (*Mapping 5*) was high for **Sample C** (focused on spinal movements), this trend does not persist for the other two samples. The standard deviation of *Mappings 1 and 2* (arm mappings) were high for **Sample A** (focused on arm movements), but again this trend does not hold for the remaining samples. All mappings had high standard deviations for **Sample C**. Although users did not choose *Mappings 1 and 2* very often for **Sample B**, the standard deviations of *Mappings 1 and 2* were much higher than other mappings that were chosen more often.

The respondents, therefore, were not purely looking for signals that had a greater oscillation magnitude or standard deviation. Further analysis is needed to determine if other aspects of the θ_x and θ_y signals correlate to the respondents' preferences to build a model of imitation preferences.

4.4.4 Discussion of Results

In this section, we will include three discussion points tied to the three hypotheses introduced in in Section 4.3.3.

Discussion of Hypothesis 1

Respondents, over all motion capture samples, preferred motion capture mappings, *Mappings 1-5*, over an artificially generated mapping, *Mapping 6*. Respondents chose, without forced response, motion capture mappings 73% of the times seen compared to only 26% for artificially generated. This result holds for individual samples as well with percentages without forced response of at least 72% for motion capture mappings. This supports **Hypothesis 1** and shows that respondents preferred motion capture motion over artificially generated motion.

Discussion of Hypothesis 2

Table 4.4: Results tabulated with forced response for each sample and mapping group combination. The entries we predicted to be the highest, where the body part for which the sample was chosen matches the mapping group, are in bold, and the highest percentage for each sample is highlighted.

	Sample A	**Sample B**	**Sample C**
Mappings 1-2	62.0%	53.3%	53.4%
Mappings 3-4	56.5%	63.7%	63.3%
Mapping 5	52.7%	59.0%	61.6%
Mapping 6	33.3%	32.8%	28.5%

We predicted from **Hypothesis 2** that a particular mapping group that tracks the body part for which a sample was chosen would have the highest preference by the users. This would be, for example, arms mappings (*Mappings 1-2*) for **Sample A** that focused on arm movements. This would mean that participants chose mappings to robot motion that reflected our observed focus of the movement in a particular motion capture sample.

Table 4.4 contains the percentages with the forced response for each sample and mapping group combination, with the predicted highest entries in bold and the highest entry for each sample highlighted. We can conclude that our predicted mapping was chosen above 50% for each sample, but was not necessarily chosen the most often. For example, legs (*Mappings 3-4*) were chosen more frequently than the spine (*Mapping 5*) for **Sample C** that focused on the spine. Looking at the percentages from Table 4.4, the predicted highest percentages for arms (*Mappings 1-2*) and legs (*Mappings 3-4*) were the highest for their respective samples.

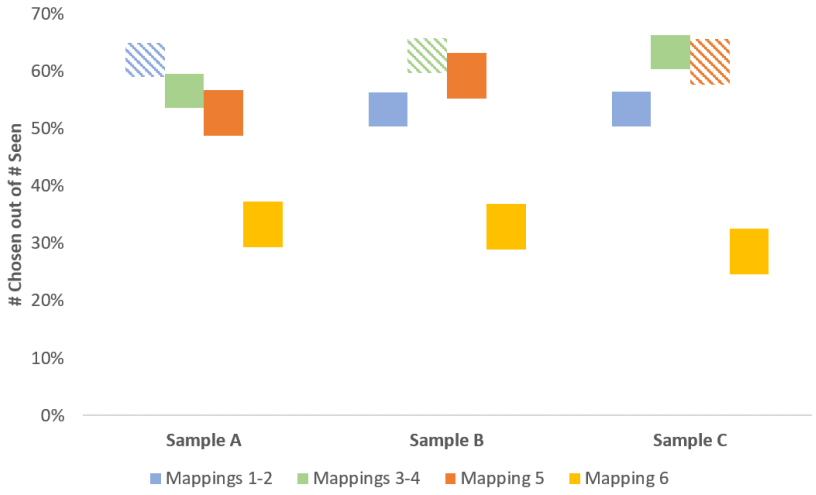


Figure 4.8: The 95% confidence interval for results with forced response, for each sample and mapping group combination. The bars indicating our predicted highest percentages, are striped.

When taking into account the 95% confidence interval for these results, this hypothesis is even less clear. We can visualize these intervals in Figure 4.8, which shows them for each mapping group separated by sample. The confidence intervals overlap for all of the mapping groups over all samples, so we cannot say definitively if our predicted mapping for each sample would be chosen more often. Although our predicted preferred mapping for each sample has a high preference (over 60%), we cannot support with our data **Hypothesis 2** that our predicted mappings for each sample will outperform other motion capture mappings.

Discussion of Hypothesis 3

From the results of previous work and insight from dancers, we predicted that verticality would be preferred. The importance of verticality was investigated with a collaborator trained in classical ballet, and the mover in the motion capture data was also trained in classical ballet. This connection led us to **Hypothesis 3**, stating the preference for verticality.

Looking at the results separated by each mapping (left column of Figure 4.5), we can see that verticality represented by *Mapping 5* was chosen most often (with forced response was 58%) over all samples. The next highest value is *Mapping 3* at 56%. However, when combining mappings into mapping groups, the verticality value remains the same, but legs (*Mappings 3-4*) has a value of 61%. This is due to some stimuli comparing *Mapping 3* and *Mapping 4*, which would always result in a choice of a mapping in the *Mappings 3-4* group.

Since this property of two robots from the same mapping group being compared occurs in the cases of

Mappings 1-2 and *Mappings 3-4*, we cannot definitively support **Hypothesis 3** with our results. Further studies are required that control for this property by either eliminating videos that compare two mappings from the same mapping group or by adding another mapping that would fit into a mapping group with *Mapping 5* to balance what the user sees.

However, when analyzing the respondent distributions in Section 4.4.2, we developed two measures, loyalty and engagement, that allowed as to make conclusions about a subgroup of the respondents. These 29 respondents chose spinal movement (*Mapping 5*) over 85% of the time, and their loyalty to verticality can be quantified using the developed loyalty measure. Additionally, we concluded that these respondents were more engaged, as it pertains to the short answer responses, when compared to the overall respondent pool.

As we saw from Figure 4.6, there were no other significant subgroups that chose any of the other three mapping groups with high frequency. This implies that, at least for these three motion capture samples, verticality was significant. Given the propensity of users of MTurk to move very quickly through surveys, the overwhelming choice of verticality by a subgroup of respondents is compelling. In further analysis, we can attempt to determine characteristics of the respondents, samples chosen, or mappings used that would explain this verticality respondent group.

Chapter 5

Human Motion Analysis for Cultural Understanding

The experimental results detailed in the previous chapter (Chapter 4) include the importance of verticality, or the leaning of the spine, for low-DOF imitation of human motion. However, as seen in Section 4.2, the motion capture sample used has an impact on how the imitation is perceived. All the recorded motion capture data from the previous chapters are of a trained ballet dancer, and that will also have an impact on how well verticality represents the human motion.

In this chapter, we discuss two Indian classical dance styles, Kathak and Bharatanatyam, and how verticality likely cannot identify the subtle differences between the two forms. We will first discuss a background of these two dance styles (Section 5.1) and then examine qualitative differences between Kathak and Bharatanatyam in Section 5.2. We will then consider how verticality would represent those differences and suggest improved measures based on our observations (Section 5.3). We conclude in Section 5.4 by describing how this analysis can lead to a greater understanding of how movements can vary as a result of context, or more specifically as a result of cultural differences, and how to leverage those differences for culturally adaptive robot motion.

5.1 Background on Kathak and Bharatanatyam

The Sanskrit text Natya Shastra (500BCE to 500CE) delves into the ancient Indian performing arts [64]. The dance section describes hand/feet positions and conveying emotions through movement and expression. The Indian National Academy for Music, Dance, and Drama recognizes eight styles of Indian classical dance - Kathak, Bharatanatyam, Kuchipudi, Kathakali, Manipuri, Odissi, Sattriya, and Mohiniyattam.

Bharatanatyam and Kathak, from southern and northern India respectively, diverged significantly from their common ancient dance ancestor due to historical, cultural and regional differences. Sharpness, tension, and straight lines in arms and legs characterize Bharatanatyam movements. In contrast, Kathak movements are softer with less tension in elbows and wrists.

Qualitative features of these dance styles have not been quantified and pose challenges to typical capture processes. How do we quantify small differences observed in similar movements from these two styles?

Figure 5.1: Verticality vector (green) with respect to positive z-axis of the mover (black). **Left:** Motion capture skeleton with angle from z-axis to verticality vector θ labelled. **Right:** A Kathak (left) and Bharatanatyam (right) dancer performing similar movements. The verticality vector does not capture differing hand gestures or tension in limbs. Screenshot from [2].

Similar research has compared other pairs of dance styles, such as Kathak and Flamenco [65].

5.2 Kathak and Bharatanatyam movement comparison

We will observe similarities and differences in an analogous position and in hand gestures performed in both dance styles. These observations were performed by the author who has trained in both the Lucknow school of Kathak and Kalakshetra school of Bharatanatyam.

5.2.1 A similar movement in two styles

Figure 5.2 shows a movement performed in both styles (left: Kathak, right: Bharatanatyam). Both dancers extend their left arm to the upper-back-left corner and point their right foot towards the bottom-front-right corner. Their right hands point inward at chest level with head turned looking at their left hand. Their bodies are angled, pointing towards the front-left. However, the Bharatanatyam dancer extends her left elbow while lunging, right knee unbent. The Kathak dancer bends her left elbow and has a more balanced stance. We therefore conclude that these dancers are performing similar movements in different styles.

The positions in Figure 5.2 also differ in hand gestures, named using [66]. The Bharatanatyam dancer's left hand is in *alapadma* (fingers splayed), and her right hand is in *katakaamukha* (first two fingers touching thumb with other two fingers splayed). The Kathak dancer's left hand is in *pataaka* (fingers outstretched

Figure 5.2: A Kathak (left) and Bharatanatyam (right) dancer performing similar movements with left arm pointing up and right foot extended out. Weight shift, tension in the limbs, and hand gestures differentiate the two positions. Screenshot from [3].

and together with thumb slightly tucked inward), while her right hand is in *araala* (index finger touching thumb with other fingers outstretched and together).

5.2.2 Hand Gesture Comparison

The same hand gesture still exhibits subtle differences when performed in these two styles. Figure 5.3 illustrates the differences in two hand gestures (*pataaka* and *araala*) performed in Kathak and Bharatanatyam. The purple circles highlight differences in thumb positioning. The Kathak gesture has lower tension in the thumb, lightly touching the side of the hand, while the Bharatanatyam gesture has higher tension in the thumb, held forcefully into the hand.

The green circles emphasize differences in muscular tension in the entire hand. In Bharatanatyam, this tension is created by arching the tightly squeezed fingers and can be observed through veins standing out prominently in the wrist. The Kathak gesture has fingers placed flatter, not pressed together as tightly, and lower tension in the wrist. These subtle differences in hand gestures exemplify the stylistic differences between Bharatanatyam and Kathak.

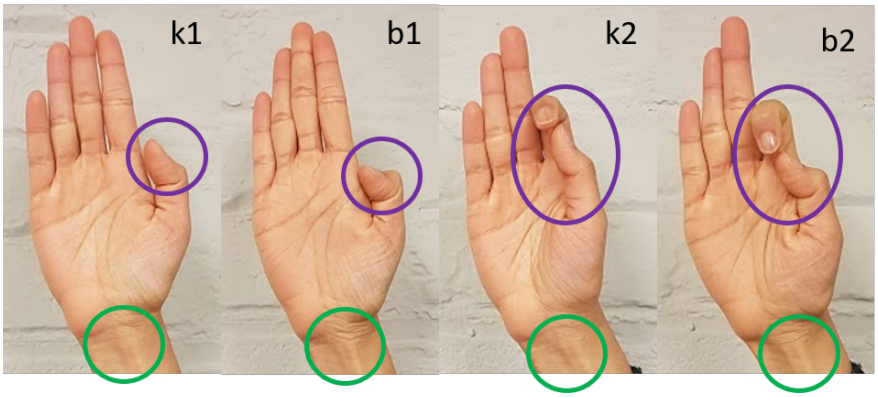


Figure 5.3: The Kathak(k) and Bharatanatyam(b) hand gestures *pataaka*(1) and *araala*(2) compared. The purple circles call attention to the thumb's positioning, and the green circles emphasize the increased tension in the wrist, spreading to the entire hand.

5.3 Verticality and other possible measures

After observing the similarities and differences in the two dance styles, we now attempt to apply verticality, discussed in previous chapters., to quantify these differences. Figure 5.1 (right) illustrates a Kathak and Bharatanatyam dancer in a similar position (mirrored). The verticality vector (green), connecting the lower neck and pelvis, indicates the spine leaning away from the z-axis (dotted black). The roughly corresponding angles formed by the two verticality vectors demonstrates the movements' similarity. However, nuances in hand gestures, limb positions, and tension cannot not captured by this measure alone.

We will propose a variety of measures to detect differences in the two dance styles not evident through verticality alone. For example, we can extend the mathematical process used to compute verticality to construct other vectors on the motion capture skeleton, such as those used in Chapter 4.3. A measure of angles made by arms and legs with the vertical may yield a richer representation of motion.

An alternative method may look at specific angles in the data set. These could include the angles between the hand and wrist, forearm and upper arm, and lower and upper leg. For example, in Figure 5.1, the elbow angle varies in the two positions because Kathak dancers tend to keep a greater bend in the elbow, to preserve the softness of the movement.

Tracking differences in tension, especially in hand gestures, may be difficult to measure. A hand motion capture system, integrated with a full body motion capture system, may be capable of tracking small differences in the hand positions between styles. However, these differences in tension may not be distinguishable through motion capture alone, requiring the use of other types of sensors.

5.4 Summary of Dance Styles Discussion

We have presented a set of observations comparing similar movements executed with different movement features (e.g. hand gestures and muscular tension) in two Indian classical dance styles. We have described limitations in a motion capture metric, verticality, to discriminate between the two styles. We have also discussed other potential measures to quantify differences in similar movements for this comparison.

Examining static positions in the two dance styles yields useful information. However, analyzing motion data sets from both dance styles where dancers perform similar movements will provide a richer quantitative comparison of Bharatanatyam and Kathak. This comparative framework can generate better motion representations valuable in a variety of applications. Verticality was useful within the context of Western dance but can break down when differentiating between two Indian dance styles. Similarly, in-home robots may need additional metrics for sensing motion in users of different cultures or environments. An understanding of these differences can allow us to develop socially adaptive robot motion that adjusts to varying cultural contexts.

Chapter 6

Conclusion

In this thesis, we first explored dyadic interactions through motion capture data to identify the importance of the leaning of spine, verticality, to the interaction between two individuals performing a task together. This investigation was inspired by the movement perspective of a dancer interested in the characteristics of partnering. We then introduced a four degree of freedom model of human motion using motion capture data to use verticality to command two low-DOF simulated robots. We ran an initial experiment testing the viability of verticality for imitation of human motion when compared to artificially generated motion on the two low-DOF simulated robots. We saw that users preferred verticality for the 4-DOF robot, Broombot, at least for the short motion capture clips from a trained ballet dancer.

We then developed and ran a study investigating the change in human perceptions of the robot imitation when we change the mapping between human and robot motion and when we change the human's motion. We found that users still preferred motion capture motion over artificially generated motion. We also found that users did not always pick the mapping to robot motion that corresponded to the area of the body we observed high deflections in the human. Lastly, we identified a sub-group of respondents who were very loyal to the verticality mapping and were more engaged in the short answer component of the survey, which may indicate that they were more attentive to the survey overall. This human preference analysis can lead to a more extensive model for human preferences for robot imitation, allowing us to determine combinations of factors in the imitation that humans are likely to prefer.

Although verticality was shown to be useful for imitation when using the human motion of a trained ballet dancer, verticality may not capture the nuances of other styles of movement. We explored the differences between two styles of Indian classical dance, Kathak and Bharatanatyam, to identify key aspects of the motion that were not explained by verticality alone. This analysis can form the foundation of developing varying motion for robots in various social and cultural situations.

6.1 In-Progress and Future Work

We are currently developing a hardware implementation of several of the concepts we have discussed in this thesis, beginning with a framework connecting motion capture to simulated robot motion in real-time. Using the built-in data streaming functionality in Motive, the motion capture software for the Optitrack system, we are building the architecture to stream the data directly into ROS. This will allow us to simulate the Rollbot and Broombot moving in real-time to a human in a motion capture suit, using the ROS software RViz. We can also extend this framework to physical hardware, such as an iRobot Create for the Rollbot and a 2-DOF inverted pendulum mounted on an iRobot Create for the Broombot. We can then develop experiments testing how the humans being imitated by the robots view various imitation mappings and move in response to the robots' motion.

6.1.1 *Menagerie* Performance

We are also currently designing a performance called *Menagerie* that showcases the tools developed in this thesis. A possible schematic of the performance space is shown in Figure 6.1. The performance includes a series of exhibits reminiscent of museum exhibits. Each exhibit will include a video with a combination of simulated robots, actual robots, humans, and motion capture skeleton. Using the code from Appendix A and the hardware framework describe in this section, a large variety of videos can be generated. Next to each video playing, there will be a card with the title of each piece and a description of the piece.

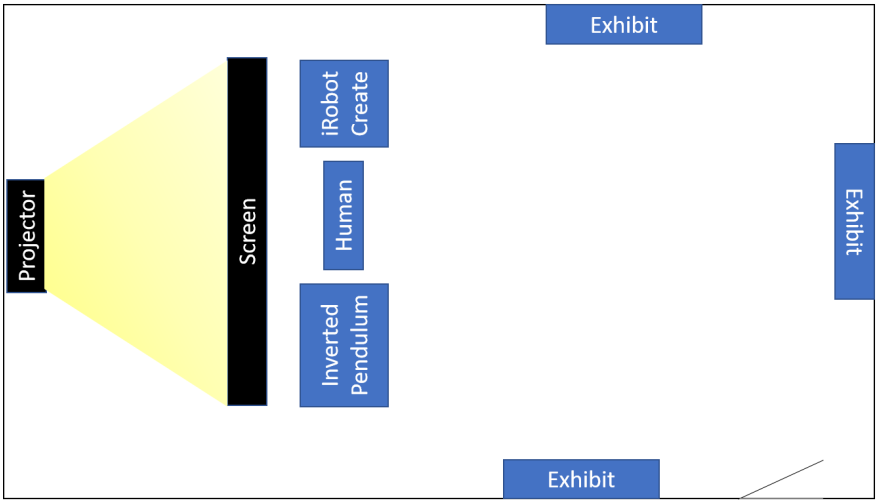


Figure 6.1: A possible schematic for the Menagerie performance space. The performance could include various museum-type exhibits that each have videos with human movers and simulated and/or actual robots. It could also include a live performance with a some subset of a human moving in front of a projector screen with simulated robots, a moving iRobot Create, and a moving inverted pendulum robot.

There is also a live performance aspect of this piece. After visitors to the piece explore the exhibits, a human can move to music in front of a projector screen that is back-projected as in Figure 6.1. Various versions of the piece could include some subset of simulated robots projected onto the screen, an iRobot Create, and an inverted pendulum robot. The performance will be designed to showcase the movements of the robots and affect how visitors view the robots' motion. The style of the mover's movements, the types of robots used, and how the movements interact with the chosen music will all affect the audience's view of the piece. Similar to previous work done in our lab, such as in [67, 68], surveys can investigate how the viewers' perceptions of the robots change after the performance.

6.2 Publications

The work done in this thesis produced the following publications.

- R. Kaushik, I. Vidrin, and A. LaViers, "Quantifying Coordination in Human Dyads via a Measure of Verticality," in Proceedings of the 5th International Conference on Movement and Computing - MOCO '18, Genoa, Italy, 2018, pp. 1–8.
- R. Kaushik and A. LaViers, "Imitating Human Movement Using a Measure of Verticality to Animate Low Degree-of-Freedom Non-humanoid Virtual Characters," in Social Robotics, Cham, 2018, vol. 11357, pp. 588–598.
- R. Kaushik and A. LaViers, "Using verticality to classify motion: analysis of two Indian classical dance styles," Symposium on "Movement that Shapes Behaviour". at Artificial Intelligence and Simulation of Behaviour (AISB) 2019. England. (to appear)
- R. Kaushik and A. LaViers, "Imitation of Human Motion by Low Degree-of-Freedom Simulated Robots and Human Preference for Mappings Driven by Spinal, Arm, and Leg Activity," International Journal of Social Robotics. (under review)

Chapter 7

References

[1] "CMU Graphics Lab Motion Capture Database." http://mocap.cs.cmu.edu.

[2] C. of the Classics, "Liquid Dance - Bharatanatyam vs. Kathak." https://www.youtube.com/watch?v=cqnSrdgmn20, 2017.

[3] D. Pandit, "Bharatanatyam-Kathak Duet." https://www.youtube.com/watch?v=p9OnKpkXHDk, 2016.

[4] M. Tomasello, A. C. Kruger, and H. H. Ratner, "Cultural learning," *Behavioral and brain sciences*, vol. 16, no. 3, pp. 495–511, 1993.

[5] C. Breazeal and B. Scassellati, "Robots that imitate humans," *Trends in cognitive sciences*, vol. 6, no. 11, pp. 481–487, 2002.

[6] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robotics and autonomous systems*, vol. 42, no. 3-4, pp. 143–166, 2003.

[7] S. M. Boker and J. L. Rotondo, "Symmetry building and symmetry breaking in synchronized movement," *Mirror neurons and the evolution of brain and language*, vol. 42, p. 163, 2002.

[8] S. M. Boker, J. F. Cohn, B.-J. Theobald, I. Matthews, T. R. Brick, and J. R. Spies, "Effects of damping head movement and facial expression in dyadic conversation using real–time facial expression tracking and synthesized avatars," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 364, no. 1535, pp. 3485–3495, 2009.

[9] K. T. Ashenfelter, S. M. Boker, J. R. Waddell, and N. Vitanov, "Spatiotemporal symmetry and multifractal structure of head movements during dyadic conversation.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 35, no. 4, p. 1072, 2009.

[10] C. Liu, C. T. Ishi, H. Ishiguro, and N. Hagita, "Generation of nodding, head tilting and eye gazing for human-robot dialogue interaction," in *Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference On*, pp. 285–292, IEEE, 2012.

[11] E. A. Mielke, E. C. Townsend, and M. D. Killpack, "Analysis of Rigid Extended Object Co-Manipulation by Human Dyads: Lateral Movement Characterization," *arXiv preprint arXiv:1702.00733*, 2017.

[12] A. Melnyk and P. Hénaff, "Physical Analysis of Handshaking Between Humans: Mutual Synchronisation and Social Context," *International Journal of Social Robotics*, Feb. 2019.

[13] K. Woolford, "Capturing human movement in the wild," in *Proceedings of the 2014 International Workshop on Movement and Computing*, p. 19, ACM, 2014.

[14] D. Arvind and A. Valtazanos, "Speckled tango dancers: Real-time motion capture of two-body interactions using on-body wireless sensor networks," in *Wearable and Implantable Body Sensor Networks, 2009. BSN 2009. Sixth International Workshop On*, pp. 312–317, IEEE, 2009.

[15] R. Kaushik, I. Vidrin, and A. LaViers, "Quantifying Coordination in Human Dyads via a Measure of Verticality," in *Proceedings of the 5th International Conference on Movement and Computing - MOCO '18*, (Genoa, Italy), pp. 1–8, ACM Press, 2018.

[16] C. Brown and G. Paine, "Interactive Tango Milonga: Designing internal experience," in *Proceedings of the 2nd International Workshop on Movement and Computing*, pp. 17–20, ACM, 2015.

[17] W. Li and P. Pasquier, "Automatic Affect Classification of Human Motion Capture Sequences in the Valence-Arousal Model," in *Proceedings of the 3rd International Symposium on Movement and Computing*, p. 15, ACM, 2016.

[18] T. Shiratori, A. Nakazawa, and K. Ikeuchi, "Synthesizing dance performance using musical and motion features," in *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.*, pp. 3654–3659, IEEE, 2006.

[19] P. Kingston and M. Egerstedt, "Motion preference learning," in *American Control Conference (ACC), 2011*, pp. 3819–3824, IEEE, 2011.

[20] M. Pomplun and M. J. Mataric, "Evaluation metrics and results of human arm movement imitation," in *Proceedings, First IEEE-RAS International Conference on Humanoid Robotics (Humanoids-2000)*, pp. 7–8, 2000.

[21] L. Noy, E. Dekel, and U. Alon, "The mirror game as a paradigm for studying the dynamics of two people improvising motion together," *Proceedings of the National Academy of Sciences*, vol. 108, no. 52, pp. 20947–20952, 2011.

[22] P. Slowinski, E. Rooke, M. Di Bernardo, and K. Tanaseva-Atanasova, "Kinematic characteristics of motion in the mirror game," in *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 748–753, IEEE, 2014.

[23] G. Maeda, M. Ewerton, R. Lioutikov, H. B. Amor, J. Peters, and G. Neumann, "Learning interaction for collaborative tasks with probabilistic movement primitives," in *Humanoid Robots (Humanoids), 2014 14th IEEE-RAS International Conference On*, pp. 527–534, IEEE, 2014.

[24] M. Ewerton, G. Neumann, R. Lioutikov, H. B. Amor, J. Peters, and G. Maeda, "Learning multiple collaborative tasks with a mixture of interaction primitives," in *Robotics and Automation (ICRA), 2015 IEEE International Conference On*, pp. 1535–1542, IEEE, 2015.

[25] S. Cuykendall, T. Schiphorst, and J. Bizzocchi, "Designing interaction categories for kinesthetic empathy: A case study of synchronous objects," in *Proceedings of the 2014 International Workshop on Movement and Computing*, p. 13, ACM, 2014.

[26] K. Özcimder, B. Dey, R. J. Lazier, D. Trueman, and N. E. Leonard, "Investigating group behavior in dance: An evolutionary dynamics approach," in *American Control Conference (ACC), 2016*, pp. 6465–6470, IEEE, 2016.

[27] K. Yamane, Y. Ariki, and J. Hodgins, "Animating non-humanoid characters with human motion data," in *Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp. 169–178, Eurographics Association, 2010.

[28] M. Abdul-Massih, I. Yoo, and B. Benes, "Motion Style Retargeting to Characters With Different Morphologies," in *Computer Graphics Forum*, vol. 36, pp. 86–99, Wiley Online Library, 2017.

[29] Y. Seol, C. O'Sullivan, and J. Lee, "Creature features: Online motion puppetry for non-human characters," in *Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp. 213–221, ACM, 2013.

[30] E. Bevacqua, R. Richard, J. Soler, and P. De Loor, "INGREDIBLE: A platform for full body interaction between human and virtual agent that improves co-presence," in *Proceedings of the 3rd International Symposium on Movement and Computing*, p. 22, ACM, 2016.

[31] J. K. Tang, J. C. Chan, and H. Leung, "Interactive dancing game with real-time recognition of continuous dance moves from 3D human motion capture," in *Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication*, p. 50, ACM, 2011.

[32] C. Zhai, F. Alderisio, P. Słowiński, K. Tsaneva-Atanasova, and M. di Bernardo, "Design of a virtual player for joint improvisation with humans in the mirror game," *PloS one*, vol. 11, no. 4, p. e0154361, 2016.

[33] C. Zhai, F. Alderisio, P. Slowinski, K. Tsaneva-Atanasova, and M. di Bernardo, "Modeling joint improvisation between human and virtual players in the mirror game," *arXiv preprint arXiv:1512.05619*, 2015.

[34] J. McCormick, K. Vincs, S. Nahavandi, D. Creighton, and S. Hutchison, "Teaching a digital performing agent: Artificial neural network and hidden markov model for recognising and performing dance movement," in *Proceedings of the 2014 International Workshop on Movement and Computing*, p. 70, ACM, 2014.

[35] M. Gillies, H. Brenton, and A. Kleinsmith, "Embodied design of full bodied interaction with virtual humans," in *Proceedings of the 2nd International Workshop on Movement and Computing*, pp. 1–8, ACM, 2015.

[36] K. Yamane, "Human Motion Tracking by Robots," in *Dance Notations and Robot Motion*, pp. 417–430, Springer, 2016.

[37] C. Ott, D. Lee, and Y. Nakamura, "Motion capture based human motion recognition and imitation by direct marker control," in *Humanoid Robots, 2008. Humanoids 2008. 8th IEEE-RAS International Conference On*, pp. 399–405, IEEE, 2008.

[38] T. Minato and H. Ishiguro, "Generating natural posture in an android by mapping human posture in three-dimensional position space," in *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference On*, pp. 609–616, IEEE, 2007.

[39] I. Fujimoto, T. Matsumoto, P. R. S. De Silva, M. Kobayashi, and M. Higashi, "Mimicking and evaluating human motion to improve the imitation skill of children with autism through a robot," *International Journal of Social Robotics*, vol. 3, no. 4, pp. 349–357, 2011.

[40] S. Nakaoka, A. Nakazawa, K. Yokoi, H. Hirukawa, and K. Ikeuchi, "Generating whole body motions for a biped humanoid robot from captured human dances," in *2003 IEEE International Conference on Robotics and Automation (Cat. No. 03CH37422)*, vol. 3, pp. 3905–3910, IEEE, 2003.

[41] Y. Demiris and M. Johnson, "Distributed, predictive perception of actions: A biologically inspired robotics architecture for imitation and learning," *Connection Science*, vol. 15, no. 4, pp. 231–243, 2003.

[42] A. Billard and M. J. Matarić, "Learning human arm movements by imitation:: Evaluation of a biologically inspired connectionist architecture," *Robotics and Autonomous Systems*, vol. 37, no. 2-3, pp. 145–160, 2001.

[43] W. Suleiman, E. Yoshida, F. Kanehiro, J.-P. Laumond, and A. Monin, "On human motion imitation by humanoid robot," in *2008 IEEE International Conference on Robotics and Automation*, pp. 2697–2704, IEEE, 2008.

[44] C. Nehaniv and K. Dautenhahn, "Mapping between dissimilar bodies: A ordances and the algebraic foundations of imitation," *EWLR-98*, pp. 64–72, 1998.

[45] A. Alissandrakis, C. L. Nehaniv, K. Dautenhahn, and J. Saunders, "Evaluation of robot imitation attempts: Comparison of the system's and the human's perspectives," in *Proceeding of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction - HRI '06*, (Salt Lake City, Utah, USA), p. 134, ACM Press, 2006.

[46] G. Van de Perre, A. De Beir, H.-L. Cao, P. G. Esteban, D. Lefeber, and B. Vanderborght, "Studying Design Aspects for Social Robots Using a Generic Gesture Method," *International Journal of Social Robotics*, Feb. 2019.

[47] R. Simmons and H. Knight, "Keep on dancing: Effects of expressive motion mimicry," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 720–727, IEEE, 2017.

[48] H. Knight and R. Simmons, "Laban head-motions convey robot state: A call for robot body language," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2881–2888, IEEE, 2016.

[49] H. Knight, "Expressive motion for low degree-of-freedom robots," 2016.

[50] M. Sharma, D. Hildebrandt, G. Newman, J. E. Young, and R. Eskicioglu, "Communicating affect via flight path: Exploring use of the laban effort system for designing affective locomotion paths," in *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction*, pp. 293–300, IEEE Press, 2013.

[51] A. D. Dragan, N. D. Ratliff, and S. S. Srinivasa, "Manipulation planning with goal sets using constrained trajectory optimization," in *2011 IEEE International Conference on Robotics and Automation*, pp. 4582–4588, IEEE, 2011.

[52] M. T. Chan, R. Gorbet, P. Beesley, and D. Kulić, "Interacting with curious agents: User experience with interactive sculptural systems," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 151–158, IEEE, 2016.

[53] H. Wang and K. Kosuge, "Control of a robot dancer for enhancing haptic human-robot interaction in waltz," *IEEE Transactions on Haptics*, vol. 5, no. 3, pp. 264–273, 2012.

[54] J. Baillieul and K. Özcimder, "The control theory of motion-based communication: Problems in teaching robots to dance," in *American Control Conference (ACC), 2012*, pp. 4319–4326, IEEE, 2012.

[55] T. Takeda, Y. Hirata, and K. Kosuge, "Dance partner robot cooperative motion generation with adjustable length of dance step stride based on physical interaction," in *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference On*, pp. 3258–3263, IEEE, 2007.

[56] J. Hölldampf, A. Peer, and M. Buss, "Synthesis of an interactive haptic dancing partner," in *RO-MAN, 2010 IEEE*, pp. 527–532, IEEE, 2010.

[57] W. Laird and J. Laird, *The Laird Technique of Latin Dancing*. International Dance Publications, 2003.

[58] C. Pallant, *Contact Improvisation: An Introduction to a Vitalizing Dance Form*. McFarland, 2006.

[59] "ROBERTO BOLLE & Greta Hodgkinson "Petite Mort"." https://www.youtube.com/watch?v=63xizexEgo, 2014.

[60] "Petite Mort by Jiri Kylian.." https://www.youtube.com/watch?v=5-_TSo6JSmo, 2015.

[61] A. Rozumalski, M. H. Schwartz, R. Wervey, A. Swanson, D. C. Dykes, and T. Novacheck, "The in vivo three-dimensional motion of the human lumbar spine during gait," *Gait & posture*, vol. 28, no. 3, pp. 378–384, 2008.

[62] M. Müller, T. Röder, M. Clausen, B. Eberhardt, B. Krüger, and A. Weber, "Documentation Mocap Database HDM05," Tech. Rep. CG-2007-2, Universität Bonn, June 2007.

[63] K. Pearson, "Notes on the history of correlation," *Biometrika*, vol. 13, no. 1, pp. 25–45, 1920.

[64] M. Ghosh *et al.*, "Natya shastra (with english translations)," *Asiatic Society of Bengal, Calcutta*, 1951.

[65] M. Phillips, "Becoming the Floor/Breaking the Floor: Experiencing the Kathak-Flamenco Connection," *Ethnomusicology*, vol. 57, no. 3, pp. 396–427, 2013.

[66] O. Bharatanatyam, "Asamyukta Hasta or Single Hand Gesture." https://onlinebharatanatyam.com/2008/01/03/asamyukta-hasta-or-single-hand-gesture/, 2008.

[67] C. Cuan, I. Pakrasi, and A. LaViers, "Time to Compile," in *Proceedings of the 5th International Conference on Movement and Computing - MOCO '18*, (Genoa, Italy), pp. 1–4, ACM Press, 2018.

[68] C. Cuan, I. Pakrasi, and A. LaViers, "Perception of Control in Artificial and Human Systems: A Study of Embodied Performance Interactions," in *Social Robotics* (S. S. Ge, J.-J. Cabibihan, M. A. Salichs, E. Broadbent, H. He, A. R. Wagner, and A. Castro-González, eds.), vol. 11357, pp. 503–512, Cham: Springer International Publishing, 2018.

Appendix A

Dancing Robot Menagerie

This document will explain step-by-step the process for obtaining data from the OptiTrack system and using the provided python code library to generate a dancing robot menagerie. The type of robot available for plotting is the Broombot

A.1 Record Data using the OptiTrack System in RAD Lab

A.1.1 Setting up the computer

1. The computer you will be using is the one by the wall of the lab
2. Make sure the Netgear box power cord is plugged in
3. Make sure the Ethernet cord from the Netgear box is plugged into the computer
4. Make sure the the USB License Key is plugged into the computer. The USB stick will be in the Motion Capture Box inside a yellow envelope
5. Make sure the USB cord from the Netgear is plugged into the computer
6. Open the Motive Software on the computer

A.1.2 Calibration

1. Open `File -> New Project` in Motive
2. Set the working directory to a folder named with your name and the date
3. Open `Layout -> Calibrate`
4. Mask visible markers for each camera after you have cleared everything from the space
5. Click `Start Wanding`
6. Take the motion capture wand and wave it around the entire capture space (don't forget to go high and low!) until you have about 2000 points for each camera
7. Click `Calculate`
8. Click `Apply`

A.1.3 Setting the Ground Plane

1. Make sure the wand is away from the space
2. Set the L-shaped piece in the center of the capture space preferably with the edges lined up with the tiles on the floor. The long side is z and the short side is x
3. Set the vertical offset to 19mm
4. Click `Calibrate`
5. Save the calibration resutls in the working directory

A.1.4 Creating the Skeleton

1. Open the skeleton pane to view the markerset. Select the Baseline Markerset.
2. Have the person who will be performing the motion put on the motion capture suit.
3. Place the markesr on the person in the same locations as the image in Motive. Make sure you rotate the skeleton to ensure you have all the markers.
4. Have the person stand in the T-position in the center of the space and click `Create`
5. Save the skeleton data

A.1.5 Recording a Take

1. Click the record button at the bottom of the screen to record.
2. Press the same button when done recording.
3. You can record as many takes as you like. They will appear on the left side labelled with a time-code.

A.1.6 Exporting the Data

1. `File -> Export`
2. Export the CSV file format with global coordinates and rotation set to XYZ. You do not need the Rigid Bodies and Rigid Body Markers selected, but make sure Markers is selected.
3. Export as a BVH file
4. Make sure you save all files (calibration, skeleton, take, csv, etc.) to your working directory before closing Motive
5. Unplug the Netgear power cord after you are done using the motion capture system and make sure all markers, suits, etc. are properly put away in the motion capture equipment box. Also make sure the USB key is put back in the yellow envelope which goes inside the motion capture equipment box.

A.2 Generating the Simulation

A.2.1 Requirements

1. Python 3 https://www.python.org/downloads/
2. Pip https://pypi.org/project/pip/

- This should work `python -m pip install -U pip`

3. Numpy and Matplotlib https://scipy.org/install.html

- This should work `python -m pip install --user numpy matplotlib`

4. FFMPEG to save the video

- How to for Windows: http://adaptivesamples.com/how-to-install-ffmpeg-on-windows/
- FFMPEG site: https://ffmpeg.zeranoe.com/builds/

5. An environment to run python (Sublime https://www.sublimetext.com/3 is nice)

A.2.2 Motion Capture Data Computations

(Only need to complete once per csv file if you are not adding any more vectors to compute!)

1. `human_data_import.py` is the file you should run. This calls the file `analyze_human_data.py`.
2. Modify the `filename` field to include the csv file from Motive, which you should move to the same folder as the repository.
3. Modify the `data_filename` field to whatever you want the `npz` file containing the computed motion capture data to be called.
4. The `dim_transfer` field adjusts the xyz coordinates to match the convention of the z-direction pointing up.
5. The next section includes a list of vectors with the start and end points as the first and second elements. If you wish to add another vector to compute, add it here and make sure you add it to the end of `vectors`
6. You should be able to run `human_data_import.py` now. It will take a little while and will print out a message when completed. The data will be saved in a `.npz` file in the same directory.

A.2.3 Video Generation

1. `main.py` is the file you want to run.
2. The options section contains the vectors from `human_data_import.py` and the `paths` to connect when plotting the human skeleton

3. The Variables to Modify section contains the options you will modify for the basic use case:
    - The positions can be renamed and added to as necessary.
    - Video Options
        - `video_filename` is the name of the `.mp4` video you want to create
        - `save_type` should be `animate` if you want to just view the animation or `video` if you want to save a video
        - `animation_speed` contains the fraction of the total points you want to use for your step size as you are animating the data. Default is 0.03.
        - `img_flag` is True if you want to save each frame as an image
    - Mover Options. The format for each mover type is:
        - Human: {`'type': 'human', 'pos': mid_pos, 'paths': paths,`
        - `... 'section': range(15000,17000), 'filename': 'mocapdata01.npz' }`
        - Broombot: {`'type': 'broombot', 'pos': right_pos, 'vector': verticality_vec,`
        - `... 'section': range(15000,17000), 'filename': 'mocapdata01.npz',...`
        - `'radius': 0.25, 'height': 0.5, 'n': 10}`
        - Rollbot: {`'type': 'rollbot', 'pos': mid_pos, 'vector': random_periodic,`
        - `... 'section': range(0,1), 'filename': 'mocapdata01.npz', 'radius': 0.25,`
        - `... 'height': 0.2, 'stretch': 1.5}`
        - Make sure all movers you want to plot are put into the list `mover_opt`
        - The `pos` field should contain one the positions from above.
        - The `vector` field should contain one of the vectors from above
        - The `section` field should contain a range of indices you want to plot. Each mover should have a the same length list here
        - The `filename` field should contain the `npz` file name saved in the same directory
        - The remaining fields modify the appearance of each mover.
4. Additional variables included:
    - `video_fps` is the frames per second of the video. 120fps is what Motive uses
    - `elevation` and `azimuth` are the perspective of the video in degrees. 15 and -180 are recommended
    - `plane_start`, `plane_end`, and `height_max` control the ground plane and z-axis limits. -5, 5, and 5 are recommended
    - `color_key` must have at least 2 elements and contains the colors scheme for the movers

A.2.4 Running the Simulation

Simply run the file `main.py` to create the simulation

Appendix B

Demographic Information for Human Preference Analysis of Imitation of Low-DOF Simulated Robots

Table B.1: Age of Participants

18-24	37
25-34	132
35-44	16
45-54	7
55-64	4

Table B.2: Gender of Participants

Male	103
Female	91
Other	0
(Blank)	2

Table B.3: Native Language of Participants

Arabic	2
Bengali	0
Chinese	1
English	149
Hindi	7
Japanese	0
Portuguese	0
Punjabi	0
Russian	0
Spanish	1
Other	39

Table B.4: Years Speaking English of Participants

Less than 1 year	0
1-2 years	1
2-5 years	3
5-10 years	18
10+	23
Native Language English	151

Table B.5: Highest Level of Education Completed of Participants

High school graduate (high school diploma or equivalent)	7
Some college but no degree	20
Associate degree in college (2-year)	7
Bachelor's degree in college (4-year)	119
Master's degree	38
Professional degree (JD, MD)	2
Doctoral degree	2
(Blank)	1

Table B.6: Experience Level of Participants in Various Movement Activities

	None	**0-2 years**	**2-4 years**	**More than 4 years**
Yoga	62	63	51	20
Running	59	51	46	40
Triathlon	98	39	30	29
Dance	70	47	62	17
Martial Arts	79	54	40	23
Sports	54	49	45	48
Swimming	84	42	41	29
Walking/ Hiking	48	50	55	43
Weight Lifting	47	41	51	57
Other	61	51	52	32