



HHS Public Access

Author manuscript

ICMI '21 Companion (2021). Author manuscript; available in PMC 2023 November 30.

Published in final edited form as:

ICMI '21 Companion (2021). 2021 October ; 2021: 362–370. doi:10.1145/3461615.3485426.

Computational Measurement of Motor Imitation and Imitative Learning Differences in Autism Spectrum Disorder:

Computational Motor Imitation Measurement in ASD

Casey J. Zampella,

Children's Hospital of Philadelphia

Evangelos Sariyanidi,

Children's Hospital of Philadelphia

Anne G. Hutchinson,

Children's Hospital of Philadelphia

G. Keith Bartley,

Children's Hospital of Philadelphia

Robert T. Schultz,

Children's Hospital of Philadelphia, University of Pennsylvania

Birkan Tunç

Children's Hospital of Philadelphia, University of Pennsylvania

Abstract

Motor imitation is a critical developmental skill area that has been strongly and specifically linked to autism spectrum disorder (ASD). However, methodological variability across studies has precluded a clear understanding of the extent and impact of imitation differences in ASD, underscoring a need for more automated, granular measurement approaches that offer greater precision and consistency. In this paper, we investigate the utility of a novel motor imitation measurement approach for accurately differentiating between youth with ASD and typically developing (TD) youth. Findings indicate that youth with ASD imitate body movements significantly differently from TD youth upon repeated administration of a brief, simple task, and that a classifier based on body coordination features derived from this task can differentiate between autistic and TD youth with 82% accuracy. Our method illustrates that group differences are driven not only by interpersonal coordination with the imitated video stimulus, but also by intrapersonal coordination. Comparison of 2D and 3D tracking shows that both approaches achieve the same classification accuracy of 82%, which is highly promising with regard to scalability for larger samples and a range of non-laboratory settings. This work adds to a rapidly growing literature highlighting the promise of computational behavior analysis for detecting and characterizing motor differences in ASD and identifying potential motor biomarkers.

Keywords

Autism; Motor imitation; Interpersonal coordination; Intrapersonal coordination; Body movement; Body joint tracking; Computational behavior analysis

1. Introduction

Motor differences in autism spectrum disorder (ASD) are highly prevalent, early-emerging, under-identified, and impactful on daily functioning [3,16,22,32]. This has led to recent calls for both routine clinical assessment of motor skills in autistic individuals and greater attention to the motor domain in autism research [4,18,34]. One barrier to the routine integration of motor measurement into clinical and research practices is reliance on clinician-administered and parent-report measures that can be resource-intensive and/or too broad to identify subtle movement patterns or change over time. Thus, development of quantitative tools for more automated and granular measurement of motor behavior across settings has the potential to significantly advance both autism research and clinical care [34].

Development of these quantitative tools may be particularly important for improving measurement of motor imitation, which has been an area of longstanding interest in ASD within the broader motor domain. Imitation plays a critical role in early development by facilitating acquisition of a range of cognitive, play, and social communication skills, leading to theories that imitation impairment may be foundational to the emergence of signature patterns of social differences in ASD (e.g., [25]). Motor imitation impairments have been steadily documented in ASD over the past several decades (see [28,31,33] for recent reviews), with meta-analysis [11] indicating significant performance gaps between individuals with and without ASD irrespective of age, gender, or movement domain (e.g., facial imitation, hand/body imitation). Imitation ability has also been significantly associated with autism symptom severity but not cognitive ability [11], suggesting that it may relate specifically to core ASD symptomatology and therefore offer a valuable target for screening, diagnosis, and treatment outcome measurement.

However, imitation studies in ASD have varied vastly in their methodology – in task parameters and in specifically how imitation performance is operationalized [28]. This variability, combined with heterogeneity within the autism spectrum itself, has precluded understanding of the precise nature and extent of imitation differences and their relationships to other functional domains [31]. For example, evidence suggests that measuring movement properties over the course of an imitated action is more meaningful for identifying differences in ASD than simply measuring the end state result [11]; however, capturing this level of detail can be difficult with traditional measurement techniques. In addition, because imitation taps both social and motor functions, there is a lack of clarity around the relative contributions of motor planning and coordination, visual perception and visuomotor integration, and social attention and motivation differences to imitation impairment in ASD [14]. Progress in addressing these unanswered questions has been hindered by the fact that imitation tasks are typically scored through human observational coding, which is both inherently subjective and staff- and time-intensive. These gaps in knowledge emphasize

the need for novel methods of quantifying imitation that are consistent, fine-grained, and scalable for widespread assessment alongside other related skill areas.

For these reasons, imitation in ASD is a ripe application for automated, granular computational behavior analysis approaches to motor measurement. Given that psychiatric conditions like ASD are defined by overt behavioral symptoms, our progress in detecting, treating, and understanding these conditions is dependent upon the accuracy and richness with which we can represent behavior. Recent innovations in computer vision and machine learning provide an opportunity to directly quantify observable human behavior from video with improved precision and consistency. Non-invasive markerless motion tracking through computer vision (e.g., [8]) offers the spatial and temporal granularity to digitize a rich range of features representing the form and kinematic properties of actions as they unfold over time (see [2] for a review in ASD). In tandem, machine learning allows for high dimensional analysis of the resulting collection of movement features to uncover patterns that best characterize differences among people with and without ASD (see [13,17]). These computational approaches are particularly well suited for imitation, as imitated motor actions are highly observable and directly quantifiable behaviors that may serve as a meaningful index of more underlying social and motor mechanisms.

Herein, we introduce a new and largely automated computational method for measuring body imitation during an explicit task in which participants are instructed to imitate a sequence of non-meaningful movements in time with a video. The specific contributions of this study are related to our comprehensive approach to imitation measurement in several novel ways:

- We track body movement across all limb joints over the full time course of the imitation task, with both a 2D (RGB) and a 3D (RGBD) camera. This allows us to examine whether 2D tracking is sufficient for uncovering group differences in how imitated movements unfold in time, or if 3D tracking provides a significant benefit.
- We incorporate information on how closely participants coordinate their body movements with the imitated video stimulus (interpersonal coordination), as well as their coordination among their own joint movements (intrapersonal coordination). This is important given pervasive gross motor coordination deficits in ASD; difficulties with coordinating one's own movements may influence the ability to coordinate with others [5,12]. This approach also offers a richer representation of the form and style with which the movements are executed.
- We measure performance across repeated administrations of a novel movement task to examine imitative motor learning. This is based on the hypothesized importance of motor learning differences in ASD [23] and recent evidence that altered imitative learning curves are significantly associated with ASD diagnostic status and symptom severity [20].

Our findings indicate that youth with ASD imitate body movements significantly differently from typically developing (TD) youth upon repeated administration of a simple 2.5-min.

task, and that a classifier that uses only body coordination features derived from this task can differentiate between ASD and TD youth with 82% accuracy. Our method illustrates that group differences are driven not only by interpersonal coordination with the imitated video stimulus, but also by intrapersonal body coordination features. Comparison of 2D vs. 3D tracking shows that both approaches achieve the same classification accuracy of 82%. The fact that a simpler, less resource-intensive method with a 2D (RGB) camera yields the same performance as an RGBD camera is highly promising with regard to scalability of data collection to large samples and a range of non-laboratory settings.

The remainder of this paper is organized as follows: Section 2 describes our method, including the sample and procedures and the calculation of inter- and intra-personal features for quantifying imitation performance and machine learning classification. Section 3 presents the results of our experiments. Section 4 discusses the significance of our results within the context of the literature on motor imitation in ASD, and Section 5 concludes the paper.

2. Methods

2.1 Sample

Participants included 39 youth (ages 9 – 17 years, 16 females) with ASD ($n = 21$) or typical development (TD; $n = 18$). All youth with ASD met Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5; [1]) criteria for ASD, confirmed in-house by an expert clinician and informed by administration of the Autism Diagnostic Observation Schedule, 2nd Edition (ADOS-2; [19]) and Social Communication Questionnaire, Lifetime Form (SCQ; [27]). For TD youth, ASD was ruled out based on the ADOS-2, SCQ, and clinician judgement of DSM-5 criteria. Participants in both groups were fluent in English and had verbal and nonverbal IQs of at least 70. ASD and TD groups were matched on age, sex, handedness, and IQ (full-scale, verbal, and nonverbal). Participant characteristics are provided in Table 1. Data collection was performed at the Center for Autism Research at Children’s Hospital of Philadelphia as part of a larger research battery examining social-motor functioning in ASD. The study was approved by the Institutional Review Board at Children’s Hospital of Philadelphia and written consent was obtained from legal guardians of all participants.

2.2 Imitation Task

The imitation task was developed based on a robust literature suggesting that sequential and non-meaningful (i.e., lacking a clear end goal) gestural imitation are particularly disrupted in ASD [33]. The task is illustrated in Figure 1a. Participants were instructed to imitate a novel sequence of gross body movements in real-time with a 2.5-min. video presented on a large TV screen in front of them. The video included six different movements performed by an adult male, with each of the six movements repeated four times in fluid succession. Some movements engaged the whole body (i.e., both arms and legs), while others only required movement of the arms with the legs remaining still. Participants were told to mirror the actions in the video stimulus, meaning that if the stimulus moved his right hand the participant should move his/her left hand. This was demonstrated with each participant by

a research assistant to ensure understanding. Each participant completed the full task twice (time-1 and time-2), separated by a brief gait measurement task, to assess the effect of imitative learning.

2.3 Body Pose Tracking

A front-facing full-body video of the participant was collected using markerless Kinect V2 cameras at a rate of 30 frames per second. Following data collection, videos were reviewed by research staff to verify that participants attended to the task and made an effort to complete it as directed. Movement at all limb joints was digitally tracked using two different approaches: (1) 2D RGB video data and the OpenPose computer vision software package [8], and (2) 3D RGBD data and the iPi Motion Capture software package [iPi Soft LLC, Moscow, Russia]. As is routinely required for the 3D motion tracking software [29], the 3D tracked skeleton was visually reviewed by staff, and the identified errors (e.g., misalignments, jitteriness) were manually corrected by refitting the tracked pose to the video. These types of tracking errors were very rare for the 2D data; post-processing was therefore not conducted in order to preserve the advantages of 2D tracking with respect to automaticity and scalability.

Tracking software yielded 13 and 20 joint coordinate sets per frame for 2D (x,y) and 3D (x,y,z) data, respectively. From these sets of joint coordinates, we analyzed movement in a subset of eight joints, namely the wrists, elbows, knees, and ankles on both the left and right sides of the body, as these were most relevant to the imitation task (see Figure 1c). This initial feature reduction was necessary for machine learning analysis given the small sample size. Each joint's distance to the body center over time was computed from coordinate data, yielding a single value per joint per frame. For each participant, this process resulted in eight time series of distance values (one per body joint) each for 2D and 3D data, at both time-1 and time-2.

2.4 Calculation of Intrapersonal and Interpersonal Joint Coordination

Imitation performance was operationalized through a set of intrapersonal and interpersonal coordination features, computed from the time series joint movement data described above. Intrapersonal coordination refers to the coordination among a person's own eight body joints in time as they execute a movement. Interpersonal coordination serves as a representation of how closely the participant coordinated with the video stimulus, by capturing the relationship between the participant's and the stimulus's joint locations in time.

Intrapersonal Coordination.—Given any two joints i, j from a participant p and a movement m , we computed the participant's intrapersonal coordination features $c_{pi pj}^m$ using windowed time-lagged cross-correlation [6]. This procedure yielded an 8x8 symmetric coordination matrix, C_p^m , as illustrated in Figure 2. The same procedure was used for the stimulus in the video (v), to compute the stimulus's intrapersonal coordination matrix C_v^m . The stimulus's matrix can be taken as the ground truth coordination matrix among the joints for each movement m . When calculating windowed time-lagged cross-correlations, we tested five different values (1s, 2s, 3s, 4s, 5s) for the window size (w), yielding five matrices, as the appropriate window size was not known a priori. The best window size was chosen

automatically within the cross-validation loop of the machine learning classifier (see Section 2.6). A lag up to a maximum value of $w/4$ was allowed in both directions to allow for small delays in the coordination between different joints.

Interpersonal Coordination.—Interpersonal coordination features were calculated similarly. We computed windowed time-lagged cross-correlations between each joint i of a participant p and the same (mirrored) joint of the video stimulus v , c_{pv}^m . We used the same five values for window size (1s, 2s, 3s, 4s, 5s), but allowed lag only in one direction with a maximum value of $w/2$, since the participant could not imitate the move before it was made in the video stimulus.

2.5 Quantifying Imitation Performance

To quantify imitation performance, we computed three sets of variables that measured: (1) the degree of similarity between the participant's intrapersonal coordination and the stimulus's (ground truth) intrapersonal coordination for each movement, (2) the degree of interpersonal coordination between the participant and the stimulus for each movement averaged across all body joints, and (3) the degree of interpersonal coordination between the participant and the stimulus for each joint averaged across all movements. These three sets of variables were computed separately for 2D and 3D data.

We calculated the first set of variables as $\alpha_p^m = 1 / \|\|C_p^m - C_v^m\|\|_*$, where $\|\|$ denotes absolute value, and $\|\|_*$ denotes the nuclear norm of a matrix. This yielded six variables, corresponding to the six movements, for each participant. For the second set of variables, we used the formula $\beta_p^m = \frac{1}{8} \sum_{i=1}^8 c_{pv}^m$, yielding another six variables. Finally, for the third set of variables, we computed eight variables, corresponding to each of the eight body joints, as $\gamma_p^j = \frac{1}{6} \sum_{m=1}^6 c_{pv}^m$. Hence, each participant's imitation performance was quantified by a set of 20 variables, capturing various aspects of imitation quality.

We also analyzed the difference in imitation performance across the two repeated administrations of the task (time-1 and time-2). We used the Wilcoxon signed-rank test to compare the variables described above across the two administrations. The effect size of those comparisons was computed using the common language effect size (CLEF; [7]) due to non-normality of the variables. CLEF, ranging between 0 and 1, is computed by testing all possible pairwise combinations of two datasets to arrive at the probability that a randomly selected score from one dataset will be higher than a randomly selected score from the other. A CLEF at either extreme of the range (i.e., near 0 or 1) is considered large, while a CLEF close to the middle of the range (i.e., near 0.5) is considered small.

2.6 Machine Learning Classification

To determine how well the set of calculated imitation variables captured differences between youth with ASD and TD youth, we used a machine learning classifier to predict the diagnostic label (ASD vs. TD) within a nested cross-validation loop. We used a support vector machine (SVM) with a linear kernel [9] as the classifier. Due to the small sample size, we used leave-one-out (LOO) cross-validation. Within each LOO iteration, using

only the training sample, the best cross-correlation window size (w) was determined via a second (inner) LOO cross-validation. Then, using the same training set, the classifier was trained, and a decision was made for the test individual. All variables were normalized to the 0–1 range, using the maximum and the minimum values learned from the training set, before being fed to the SVM classifier. We chose a linear machine learning model for prediction because of its superiority with respect to the interpretability of results, as it yields interpretable feature importance maps. We implemented the machine learning pipeline using the Scikit-learn library [24] in the Python programming language, with the default settings for a linear SVM classifier.

3. Results

The above methods provided four different datasets of 20 variables for each participant, across two digital capture technologies (2D and 3D) and two administrations (time-1 and time-2). Below, we focus on the results from the 3D dataset at time-2 and make relevant comparisons between the four different datasets. Classification performance for all datasets is presented in Table 2.

3.1 Intra- and Interpersonal Coordination and Imitation Performance

Figure 2 illustrates intrapersonal coordination matrices for the video stimulus (C_v^m) and two participants (C_p^m), one with the highest α_p^m values (high intrapersonal imitation) and another with the lowest α_p^m values (low intrapersonal imitation). Participants with high and low imitation can be distinguished by the similarity between their matrices and the matrix of the video stimulus (α_p^m). As seen in Figure 2, different patterns of intrapersonal coordination also distinguish between the six different movements that make up the task.

For all movements, there was a significant ($p < 0.05$) correlation between intrapersonal coordination similarity, α_p^m , and overall interpersonal coordination, β_p^m (move-1: $r = 0.32$, $p = 5 \times 10^{-2}$; move-2: $r = 0.49$, $p = 2 \times 10^{-3}$; move-3: $r = -0.34$, $p = 4 \times 10^{-2}$; move-4: $r = 0.64$, $p = 1 \times 10^{-5}$; move-5: $r = 0.73$, $p = 1 \times 10^{-7}$; move-6: $r = 0.65$, $p = 8 \times 10^{-6}$). The significant yet imperfect correlations suggest that these two variable sets for operationalizing imitation performance may provide complementary information regarding how successful a participant was in imitating the video stimulus's movements.

3.2 Machine Learning Classification

The machine learning classifier achieved high cross-validation accuracy in classifying ASD vs. TD using the set of 20 variables capturing intrapersonal and interpersonal coordination; Classification Accuracy = 0.82, Sensitivity = 0.76, Specificity = 0.89, Positive Predictive Value (PPV) = 0.89, Negative Predictive Value (NPV) = 0.76 (see Table 2).

The window size (w) selection for windowed cross-correlation was extremely stable. Within every cross-validation fold, $w = 3$ was selected with no exception. Similarly, feature weights were very stable between the folds, suggesting that the classifier was able to learn how ASD and TD groups were separated within our sample. Figure 3 illustrates the contribution of each feature (i.e., feature weights) in classifying participants. The highest (absolute) weight

was assigned to interpersonal coordination during movement 3, followed by intrapersonal coordination similarity between video stimulus and participant during movement 1. In terms of individual body joints, interpersonal coordination for the right wrist, left wrist, and left elbow had the highest (absolute) weights.

3.3 2 D vs. 3D Pose Tracking—We compared classification results based on the two different body tracking technologies (3D depth data and 2D RGB data), with notably similar results. More specifically, when using OpenPose over 2D images, the classification performance at time-2 was comparable to that of using 3D depth data, with Classification Accuracy = 0.82, Sensitivity = 0.81, Specificity = 0.83, PPV = 0.85, and NPV = 0.79 (see Table 2).

3.4 Learning Effects

Each participant performed the task (all six movements) twice (time-1 and time-2) to assess whether the effect of repeated administration (i.e., imitative motor learning) differed among the ASD and TD groups. We compared imitation performance across the two administrations in terms of the individual variables by univariate testing (Wilcoxon signed-rank test), as well as multivariate interactions among variables by comparing machine learning classification performance. For this purpose, we ran the machine learning classifier twice using variables from the two different administrations.

The individual variables (α^m , β^m , γ^i), when analyzed for 3D data across the entire sample, did not show any statistically significant differences between administrations 1 and 2. In the TD group alone, several variables showed indication of significant improvement (higher coordination among participant and video stimulus) at administration 2. Interpersonal coordination over all joints (β^m) for movement 6 showed significant improvement (CLEF = 0.67, $p = 2 \times 10^{-3}$), as did interpersonal coordination across all movements for the right wrist (CLEF = 0.68, $p = 2 \times 10^{-3}$) and the right ankle (CLEF = 0.60, $p = 3 \times 10^{-2}$). No significant improvements between administrations 1 and 2 were found in the ASD group. Note that none of those differences would survive a multiple comparison correction for statistical significance.

The machine learning classifiers also reflected a differential effect of imitative learning across the ASD and TD groups through lower classification accuracy at administration 1 relative to administration 2 (see Table 2). The classification performance using 3D data was significantly lower at the first administration, with Classification Accuracy = 0.62, Sensitivity = 0.62, Specificity = 0.61, PPV = 0.65, NPV = 0.58. The same was true for the 2D data, with Classification Accuracy = 0.51, Sensitivity = 0.62, Specificity = 0.39, PPV = 0.54, NPV = 0.47. The stronger differences between time-1 and time-2 reflected through the machine learning analyses relative to the univariate analyses may be because classifiers (unlike univariate testing) consider multivariate interactions across movements and joints at the participant level.

4. Discussion

Motor imitation is a critical developmental skill area that has been strongly and specifically linked to the presence and severity of ASD [11]. In this paper, we present a quantitative computational approach for measuring motor imitation that provides a degree of spatial and temporal granularity not achievable with traditional observational tools. This granularity is particularly important for increasing understanding of imitation in ASD, as evidence indicates that a specific difficulty for autistic individuals is imitating the style or form of an action as opposed to the endpoint alone [11], and because more granular measurement allows for more precise examination of key relationships between imitation and other skill areas. Our approach is also novel in its comprehensiveness, in that we quantify imitation through a set of both intrapersonal and interpersonal coordination features, across repeated task administrations, and across 2D and 3D motion tracking systems.

The results of our machine learning classification analysis are consistent with a large existing literature documenting the presence of imitation performance differences between autistic and TD youth. A machine learning classifier taking both intrapersonal and interpersonal coordination features as input was able to classify between groups with high (up to 82%) accuracy. These findings join a small group of previous studies that have applied machine learning to motor imitation features in children and adults with similar classification accuracies [17,29,30]. Collectively, these studies highlight motor imitation as one skill area in which youth with ASD can be clearly distinguished from their neurotypical peers, outside of the core ASD symptom domains of social communication and interaction and restricted/repetitive behavior. These findings have implications for screening and detection, suggesting that performance on a brief, simple imitation task may provide useful information for identifying people with ASD from those without. These results are also timely in that motor behaviors are currently receiving increased attention for their potential as biomarkers, given that they are more directly observable and quantifiable than many other psychological constructs (e.g., social cognition).

Classification accuracy between ASD and TD groups was higher for the second administration of the imitation task compared to the first, and several individual variables showed significant improvement between time-1 and time-2 in the TD group only. This suggests that TD youth may benefit more than autistic youth from opportunities for imitative learning. That is, it is possible that only the TD group's performance improved significantly with learning, leading to greater gaps between ASD and TD group performance at time-2. This was not explicitly tested and should be examined further in future analyses; however, it is consistent with work by McAuliffe and colleagues [20] showing that altered learning curves over repeated administrations of a gestural imitation task were significantly associated with ASD diagnostic status and symptom severity. The authors found that TD children plateaued in their learning early, while children with ASD required four repetitions to converge with their TD peers' performance. Thus, it appears important to consider the role of imitative learning when interpreting motor imitation and coordination performance in ASD [23].

Our classification results are remarkably consistent across datasets from an open source 2D tracking software (OpenPose; [8]) and the more established iPi Motion Capture 3D system [iPi Soft LLC, Moscow, Russia]. This result is unexpected as 3D body pose tracking is considered the gold standard. 3D tracking is also more resource-intensive in that it requires specialized cameras, expensive software, and a fair amount of human post-processing. Achieving similar results with a freely available, fully automated 2D joint tracking system that can be used with standard HD video is highly promising in terms of scalability for large samples, repeated assessment, and deployment to a range of non-laboratory settings. Providing a direct comparison of our method's performance across 2D and 3D datasets is a considerable strength of this study and expands upon similar previous work that has only analyzed 3D joint data [29]. Notably, the largest differences between the 2D and 3D datasets were for lower leg joints during imitated actions that did not engage the legs (particularly movements 1 and 2), suggesting that OpenPose may generate some systematic but spurious correlations when limbs are generally still.

Another strength of our approach is its integration of both individual-level (intrapersonal) and dyadic (interpersonal) coordination features in quantifying imitation performance. The degree to which motor imitation differences in ASD are driven by pervasive general motor coordination differences [12] remains uncertain, but there is evidence to suggest that generalized motor impairment explains some, but not all, of the variance in motor imitation ability in ASD [31]. While we did not explicitly test this by including a measure of general motor skill, our inclusion of intrapersonal coordination features accounts for the role that within-person motor coordination may play in the ability to imitate others [5]. Moreover, intrapersonal coordination features may capture the collective form of a participant's body movement as it unfolds over time better than interpersonal coordination features. Inspection of the feature weights assigned by the machine learning classifier suggests that both intrapersonal and interpersonal coordination features are important for group classification, and may have differential importance across the six different movements (Figure 3). Representing imitation in terms of both how smoothly an individual's body joints move collectively in time, as well as how closely they match a stimulus, may elucidate subtle, fine-grained patterns that contribute to observed imitation differences. Of note, regarding the feature weights for individual body joints, it is unsurprising that the two wrists were assigned the highest weights (Figure 3), as all six movements within the task required lower arm movement while only three of the movements engaged the legs.

Specific characteristics of this study's sample and methods should be kept in mind when considering the findings. First, the sample size is small and participants with ASD have average to above average language and cognitive ability. All participants were able to attend to and participate in the full task. For these reasons, results are not likely representative of or generalizable to the autism spectrum as a whole. Second, considering how imitation is defined, operationalized, and elicited is critical for situating results of individual studies within the larger context of imitation research in ASD [28]. We used a non-meaningful, sequential, explicit (i.e., not spontaneous) body imitation task, in which successful performance required fidelity to the full form of the movement rather than just an end state. Research suggests that distinct types of imitation may be differentially impacted in ASD, rely on different underlying mechanisms, and show different developmental patterns

[28, 31]. Third, we calculated coordination features based only on joints' distances from the body's center over time, without respect to directionality. The lack of information regarding angles and direction may be contributing to the similarity of results across the 2D and 3D datasets, and we might expect 3D to outperform 2D when also deriving angles. Performance metrics that take the direction of movements into account should be calculated in a future study.

It is also important to note that the focus of this study was on whether our novel computational approach could differentiate between imitation performance in autistic and TD youth. The method presented herein does not quantify imitation performance in terms of an interpretable, continuous score. It also does not directly compare our approach to existing human observational or clinician-administered methods. Related work by Tunçgenç and colleagues [29] used a similar imitation task to produce a continuous imitation performance score, which showed a strong correlation with core ASD symptom severity. These authors also compared their measure against human observational coding and found that the computational measure better distinguished between ASD and TD groups. Replicating their findings with our method is a key future direction, to establish its validity, facilitate investigation of relationships between imitation and other functional domains, and allow for tracking of imitation skill over time (e.g., as a marker of treatment outcomes). Imitation has been found to be a valuable target for intervention in ASD, with improvements seen in both imitation skill itself and broader social skills (e.g., [15]). Thus, the availability of valid, fine-grained, scalable imitation measures has the potential to be of significant value to intervention work in ASD.

5. Conclusions

In this paper, we investigated the utility of a novel quantitative motor imitation measurement approach for accurately differentiating between youth with autism spectrum disorder (ASD) and typically developing (TD) youth. Our approach rests on computational behavior analysis of a brief, simple motor imitation task administered twice. We quantified imitation performance through a combination of intrapersonal and interpersonal coordination features, computed from both 2D and 3D body joint tracking data. Our results indicate that these intra- and interpersonal coordination features can classify between autistic and TD youth with up to 82% accuracy. The highest accuracy was achieved on the second administration of the task, suggesting group differences in imitative learning. Classification accuracy was consistent regardless of whether the classifier used features extracted from 2D or 3D tracking. This work adds to a rapidly growing literature highlighting the promise of computational behavior analysis for detecting and characterizing motor differences and identifying potential motor biomarkers in ASD (e.g., see [10]). Direct acquisition and fine-grained analysis of movement data from video provides a precise, scalable tool for measuring imitation performance, particularly with respect to change over time, individual variability, and associations with other skill domains and underlying mechanisms. Important directions for future work include: (1) testing the validity of this novel approach by comparing it against established imitation metrics, (2) developing a continuous imitation score, in order to directly measure change and investigate relationships with other constructs, and (3) expanding our computational methods to a range of samples and

imitation tasks. An additional exciting future application of this work is expansion to more spontaneous and naturalistic imitation contexts, as spontaneous interpersonal coordination within social settings is a known area of difficulty for many autistic individuals [21, 26, 35, 36].

Acknowledgments

This work is partially funded by the Office of the Director (OD), National Institutes of Health (NIH), and the National Institute of Mental Health (NIMH) under grants R01MH118327 and R01MH122599.

REFERENCES

- [1]. American Psychiatric Association. 2013. Diagnostic and Statistical Manual of Mental Disorders (5th ed.). American Psychiatric Association. DOI:10.1176/appi.books.9780890425596
- [2]. de Belen Ryan Anthony J., Bednarz Tomasz, Sowmya Arcot, and Del Favero Dennis. 2020. Computer vision in autism spectrum disorder research: a systematic review of published studies from 2009 to 2019. *Translational Psychiatry* 10, 1–20. DOI:10.1038/s41398-020-01015-w [PubMed: 32066695]
- [3]. Bhat Anjana N. 2021. Motor impairment increases in children with autism spectrum disorder as a function of social communication, cognitive and functional impairment, repetitive behavior severity, and comorbid diagnoses: a SPARK study report. *Autism Res.* 14, 1 (January 2021), 202–219. DOI:10.1002/aur.2453 [PubMed: 33300285]
- [4]. Bhat Anjana Narayan. 2020. Is motor impairment in autism spectrum disorder distinct from developmental coordination disorder a report from the SPARK study. *Phys. Ther* 100, 4 (April 2020), 633–644. DOI:10.1093/ptj/pzz190 [PubMed: 32154876]
- [5]. Bloch Carola, Vogeley Kai, Georgescu Alexandra L., and Falter-Wagner Christine M. 2019. Intrapersonal synchrony as constituent of interpersonal synchrony and its relevance for autism spectrum disorder. *Front. Robot. AI* 6, (August 2019), 73. DOI:10.3389/frobt.2019.00073 [PubMed: 33501088]
- [6]. Boker Steven M., Xu Minquan, Rotondo Jennifer L., and King Kadijah. 2002. Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series. *Psychol. Methods* 7, 3 (2002), 338–355. DOI:10.1037/1082-989X.7.3.338 [PubMed: 12243305]
- [7]. Brooks Margaret E., Dalal Dev K., and Nolan Kevin P. 2014. Are common language effect sizes easier to understand than traditional effect sizes. *J. Appl. Psychol* 99, 2 (March 2014), 332–340. DOI:10.1037/a0034745 [PubMed: 24188393]
- [8]. Cao Zhe, Simon Tomas, Wei Shih En, and Sheikh Yaser. 2017. Realtime multi-person 2D pose estimation using part affinity fields. In *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, Institute of Electrical and Electronics Engineers Inc., 1302–1310. DOI:10.1109/CVPR.2017.143
- [9]. Cortes Corinna and Vapnik Vladimir. 1995. Support-vector networks. *Mach. Learn* 20, 3 (September 1995), 273–297. DOI:10.1007/bf00994018
- [10]. Dawson Geraldine and Sapiro Guillermo. 2019. Potential for digital behavioral measurement tools to transform the detection and diagnosis of autism spectrum disorder. *JAMA Pediatrics* 173, 305–306. DOI:10.1001/jamapediatrics.2018.5269 [PubMed: 30715131]
- [11]. Edwards Laura A. 2014. A meta-analysis of imitation abilities in individuals with autism spectrum disorders. *Autism Res.* 7, 3 (2014), 363–380. DOI:10.1002/aur.1379 [PubMed: 24863681]
- [12]. Fournier Kimberly A., Hass Chris J., Naik Sagar K., Lodha Neha, and Cauraugh James H. 2010. Motor coordination in autism spectrum disorders: a synthesis and meta-analysis. *J. Autism Dev. Disord* 40, 10 (October 2010), 1227–1240. DOI:10.1007/s10803-010-0981-3 [PubMed: 20195737]
- [13]. Georgescu Alexandra Livia, Koehler Jana Christina, Weiske Johanna, Vogeley Kai, Koutsouleris Nikolaos, and Falter-Wagner Christine. 2019. Machine learning to study social interaction

- difficulties in ASD. *Front. Robot. AI* 6, (November 2019), 132. DOI:10.3389/frobt.2019.00132 [PubMed: 33501147]
- [14]. Gowen Emma. 2012. Imitation in autism: Why action kinematics matter. *Front. Integr. Neurosci* 6, DEC (December 2012). DOI:10.3389/fnint.2012.00117
- [15]. Ingersoll Brooke. 2012. Brief report: Effect of a focused imitation intervention on social functioning in children with autism. *J. Autism Dev. Disord* 42, 8 (August 2012), 1768–1773. DOI:10.1007/s10803-011-1423-6 [PubMed: 22146934]
- [16]. Ketcheson Leah R., Pitchford E. Andrew, and Wentz Chandler F. 2021. The relationship between developmental coordination disorder and concurrent deficits in social communication and repetitive behaviors among children with autism spectrum disorder. *Autism Res.* 14, 4 (April 2021), 804–816. DOI:10.1002/aur.2469 [PubMed: 33421296]
- [17]. Li Baihua, Sharma Arjun, Meng James, Purushwalkam Senthil, and Gowen Emma. 2017. Applying machine learning to identify autistic adults using imitation: an exploratory study. *PLoS One* 12, 8 (August 2017), e0182652. DOI:10.1371/journal.pone.0182652
- [18]. Licari Melissa K, Alvares Gail A, Varcin Kandice, Evans Kiah L, Cleary Dominique, Reid Siobhan L, Glasson Emma J, Bebbington Keely, Reynolds Jess E, Wray John, and Whitehouse Andrew J O. 2019. Prevalence of motor difficulties in autism spectrum disorder: analysis of a population-based cohort. *Autism Res.* 13, (2019), 298–306. DOI:10.1002/aur.2230 [PubMed: 31625694]
- [19]. Lord C, Luyster R, Gotham K, and Guthrie W. 2012. *Autism Diagnostic Observation Schedules* (2nd ed.). Western Psychological Services. Retrieved June 23, 2021 from <https://eprovide.mapitrust.org/instruments/autism-diagnostic-observation-schedule>
- [20]. McAuliffe Danielle, Zhao Yi, Pillai Ajay S., Ament Katarina, Adamek Jack, Caffo Brian S., Mostofsky Stewart H., and Ewen Joshua B. 2020. Learning of skilled movements via imitation in ASD. *Autism Res.* 13, 5 (May 2020), 777–784. DOI:10.1002/aur.2253 [PubMed: 31876983]
- [21]. McNaughton Kathryn A. and Redcay Elizabeth. 2020. Interpersonal synchrony in autism. *Curr. Psychiatry Rep* 22, 3 (March 2020), 12. DOI:10.1007/s11920-020-1135-8 [PubMed: 32025922]
- [22]. Miller Haylie L., Sherrod Gabriela M., Mauk Joyce E., Fears Nicholas E., Hynan Linda S., and Tamplain Priscila M. 2021. Shared features or co-occurrence? evaluating symptoms of developmental coordination disorder in children and adolescents with autism spectrum disorder. *J. Autism Dev. Disord* (January 2021), 1–13. DOI:10.1007/s10803-020-04766-z
- [23]. Mostofsky Stewart H., Dubey Prachi, Jerath Vandna K., Jansiewicz Eva M., Goldberg Melissa C., and Denckla Martha B. 2006. Developmental dyspraxia is not limited to imitation in children with autism spectrum disorders. *J. Int. Neuropsychol. Soc* 12, 3 (May 2006), 314–326. DOI:10.1017/S1355617706060437 [PubMed: 16903124]
- [24]. Pedregosa Fabian, Varoquaux Gaël, Gramfort Alexandre, Vincent Michel, and Thirion Bertrand. 2011. Scikit-learn: machine learning in python. *J Mach Learn Res.* 12, 2825–2830.
- [25]. Rogers Sally J. and Pennington Bruce F. 1991. A theoretical approach to the deficits in infantile autism. *Dev. Psychopathol* 3, 2 (1991), 137–162. DOI:10.1017/S0954579400000043
- [26]. Romero Veronica, Fitzpatrick Paula, Roulier Stephanie, Duncan Amie, Richardson Michael J., and Schmidt RC 2018. Evidence of embodied social competence during conversation in high functioning children with autism spectrum disorder. *PLoS One* 13, 3 (March 2018). DOI:10.1371/journal.pone.0193906
- [27]. Rutter M, Bailey A, and Lord C. 2003. *Social communication questionnaire (SCQ)*. Western Psychological Services. Retrieved June 23, 2021 from https://scholar.google.com/scholar_lookup?title=Thesocialcommunicationquestionnaire&author=M.Rutter&author=A.Bailey&author=C.Lord&publication_year=2003
- [28]. Sevlever Melina and Gillis Jennifer M. 2010. An examination of the state of imitation research in children with autism: issues of definition and methodology. *Research in Developmental Disabilities* 31, 976–984. DOI:10.1016/j.ridd.2010.04.014 [PubMed: 20471220]
- [29]. Tunçgenç Bahar, Pacheco Carolina, Rochowiak Rebecca, Nicholas Rosemary, Rengarajan Sundararaman, Zou Erin, Messenger Brice, Vidal René, and Mostofsky Stewart H. 2021. Computerized Assessment of Motor Imitation as a Scalable Method for Distinguishing Children

- With Autism. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 6, 3 (March 2021), 321–328. DOI:10.1016/j.bpsc.2020.09.001 [PubMed: 33229247]
- [30]. Vabalas Andrius, Gowen Emma, Poliakoff Ellen, and Casson Alexander J. 2020. Applying machine learning to kinematic and eye movement features of a movement imitation task to predict autism diagnosis. *Sci. Rep* 10, 1 (December 2020), 1–13. DOI:10.1038/s41598-020-65384-4 [PubMed: 31913322]
- [31]. Vanvuchelen Marleen, Roeyers Herbert, and Weerdts Willy De. 2011. Do imitation problems reflect a core characteristic in autism? Evidence from a literature review. *Research in Autism Spectrum Disorders* 5, 89–95. DOI:10.1016/j.rasd.2010.07.010
- [32]. West Kelsey L. 2019. Infant motor development in autism spectrum disorder: a synthesis and meta-analysis. *Child Dev.* 90, 6 (November 2019), 2053–2070. DOI:10.1111/cdev.13086 [PubMed: 29766496]
- [33]. Williams Justin H.G., Whiten Andrew, and Singh Tulika. 2004. A systematic review of action imitation in autistic spectrum disorder. *Journal of Autism and Developmental Disorders* 34, 285–299. DOI:10.1023/B:JADD.0000029551.56735.3a [PubMed: 15264497]
- [34]. Wilson Rujuta B., McCracken James T., Rinehart Nicole J., and Jeste Shafali S. 2018. What’s missing in autism spectrum disorder motor assessments? *Journal of Neurodevelopmental Disorders* 10. DOI:10.1186/s11689-018-9257-6
- [35]. Zampella Casey J., Bennetto Loisa, and Herrington John D. 2020. Computer vision analysis of reduced interpersonal affect coordination in youth with autism spectrum disorder. *Autism Res.* 13, 12 (December 2020), 2133–2142. DOI:10.1002/aur.2334 [PubMed: 32666690]
- [36]. Zampella Casey J, Csumitta Kelsey D, Simon Emily, and Bennetto Loisa. 2020. Interactional synchrony and its association with social and communication ability in children with and without autism spectrum disorder. *J. Autism Dev. Disord* 50, (2020), 3195–3206. DOI:10.1007/s10803-020-04412-8 [PubMed: 32065341]

CCS CONCEPTS • Computer vision • Activity recognition and understanding • Supervised learning by classification

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

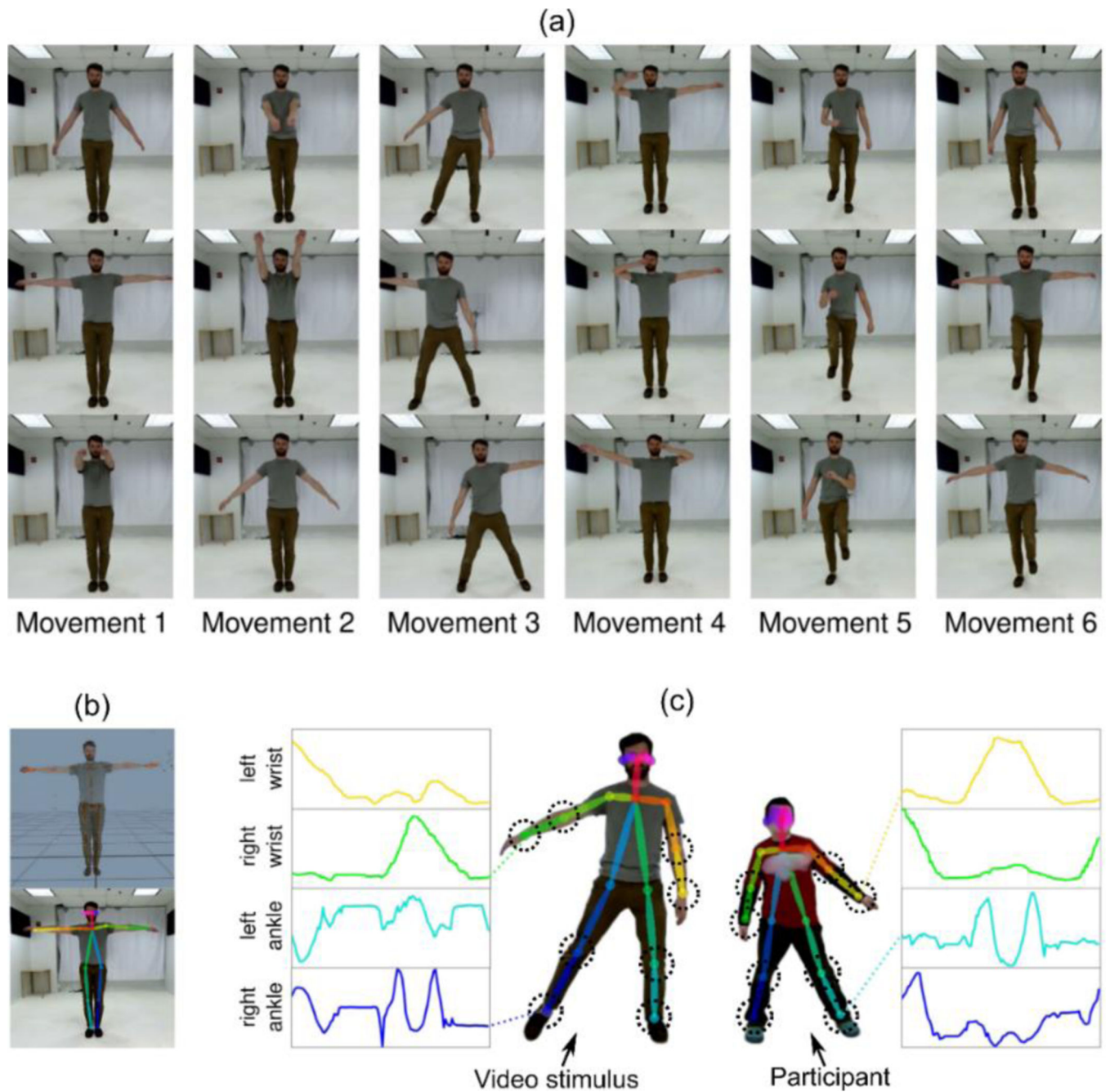
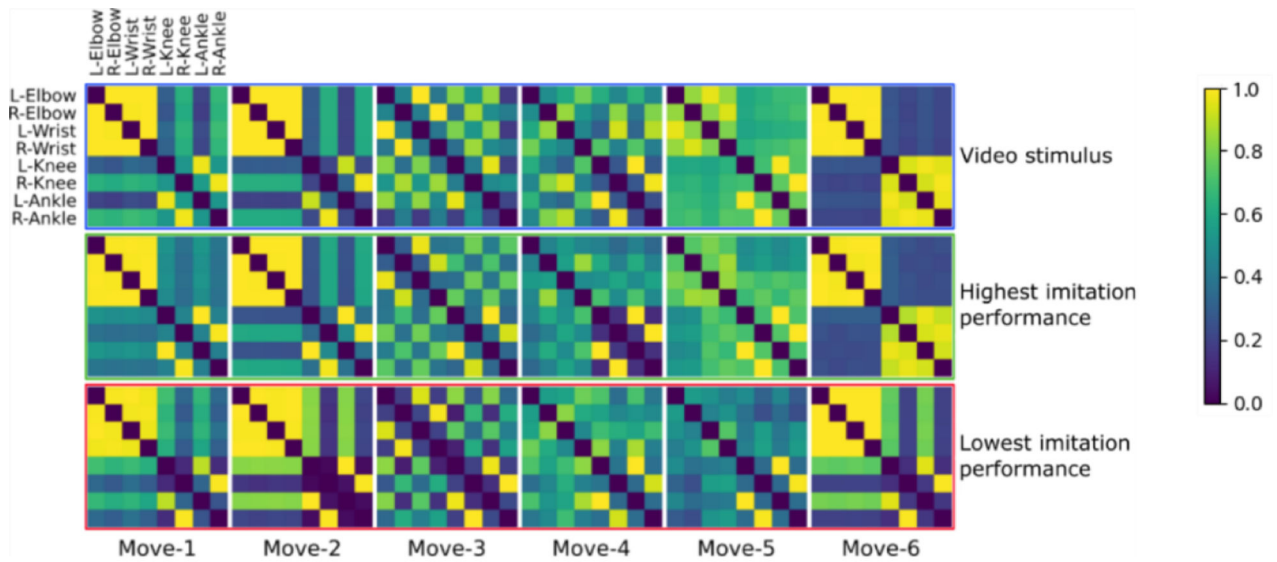
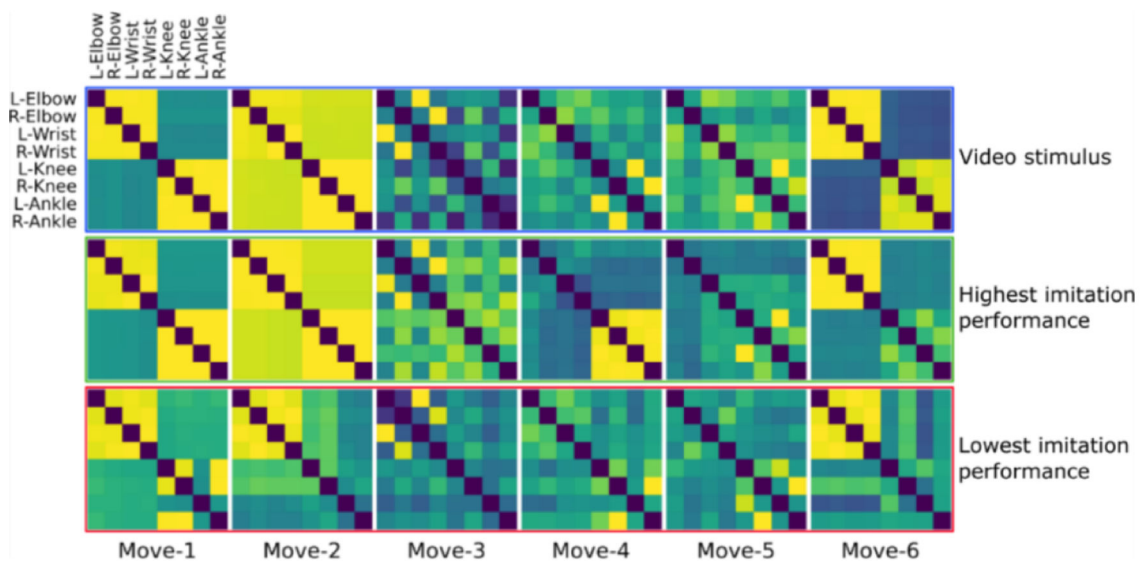


Figure 1.

Imitation task and methods (a) The task is comprised of six movements. (b) The body joint tracking methods used were 3D tracking with Kinect (top) and 2D tracking with OpenPose (bottom). (c) The imitation task in action; participants imitated the movements in the video stimulus and we used eight joints (circled) when quantifying imitation.



(a) 3D Data using depth sensors



(b) 2D Data using RGB sensors

Figure 2.

Intrapersonal coordination matrices among joints (C_j^m) for the video stimulus and two participants, presented for each movement and for (a) 3D and (b) 2D body tracking data. The first row presents matrices for the stimulus, the second row for the participant with the highest similarity to the stimulus, and the third row for the participant with the lowest similarity to the stimulus. 3D and 2D data generally resulted in similar intrapersonal coordination matrices, though matrices were less similar for movements 1 and 2. Colors represent the degree of correlation among joints.

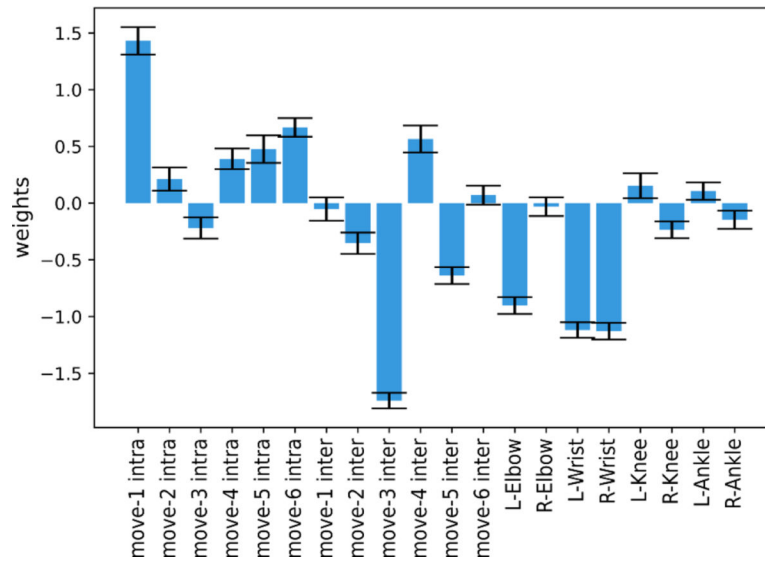


Figure 3.

Weights of individual variables as assigned by the machine learning classifier. The first six variables quantify similarity between the intrapersonal coordination of the participant and the video stimulus for each movement (α_p^m). The second set of six variables quantify the interpersonal coordination between the participant and the stimulus across all joints for each movement (β_p^m). The final eight variables quantify the interpersonal coordination for each body joint between the participant and the stimulus across all movements (γ_p^j). The higher the absolute value of a weight, the more the variable contributes to the classifier.

Table 1.

Participant characteristics

	ASD	TD	<i>t</i> or χ^2	<i>p</i>
n	21	18		
Sex, Male:Female	12:9	11:7	0.06	.80
Handedness, Right:Left	19:2	16:2	0.03	.87
Age in years, <i>M</i> (<i>SD</i>)	12.5 (2.9)	12.6 (2.3)	-0.17	.87
Full Scale IQ, <i>M</i> (<i>SD</i>)	106.2 (17.1)	103.9 (12.0)	0.44	.66
Verbal IQ, <i>M</i> (<i>SD</i>)	105.8 (15.5)	102.1 (12.7)	0.76	.46
Nonverbal IQ, <i>M</i> (<i>SD</i>)	104.2 (18.2)	105.1 (12.6)	-0.17	.86
ADOS-2 Calibrated Severity Score, <i>M</i> (<i>SD</i>)	6.9 (2.0)	1.1 (0.4)	10.90	< .001

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2.

Machine learning classification performance for 3D and 2D data at both task administrations. PPV: Positive predictive value. NPV: Negative predictive value.

Data Type	Administration	Accuracy	Sensitivity	Specificity	PPV	NPV
3D	1	0.62	0.62	0.61	0.65	0.58
3D	2	0.82	0.76	0.89	0.89	0.76
2D	1	0.51	0.62	0.39	0.54	0.47
2D	2	0.82	0.81	0.83	0.85	0.79

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript