Research in Autism Spectrum Disorders 87 (2021) 101840

Contents lists available at ScienceDirect

Research in Autism Spectrum Disorders

journal homepage: www.elsevier.com/locate/rasd

Automated and scalable Computerized Assessment of Motor Imitation (CAMI) in children with Autism Spectrum Disorder using a single 2D camera: A pilot study

Daniel E. Lidstone [a,b,*], Rebecca Rochowiak [a], Carolina Pacheco [c,d], Bahar Tunçgenç [e], Rene Vidal [c,d], Stewart H. Mostofsky [a,b,f]

[a] Center for Neurodevelopmental and Imaging Research, Kennedy Krieger Institute, Baltimore, MD, 21205, USA
[b] Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, 21205, USA
[c] Mathematical Institute for Data Science, Johns Hopkins University, 3400 N Charles St., Baltimore, MD, 21218, USA
[d] Department of Biomedical Engineering, Johns Hopkins University, 3400 N Charles St., Baltimore, MD, 21218, USA
[e] School of Psychology, University of Nottingham, University Park, Nottingham, NG7 2RD, UK
[f] Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, 21205, USA

ARTICLE INFO

Number of reviews completed is 2

Keywords:
Autism
Imitation
Intervention
Social behavior
OpenPose

ABSTRACT

Background: Motor imitation difficulties are pervasive in children with Autism Spectrum Disorder (ASD). Previous research demonstrated the validity and reliability of an algorithm called Computerized Assessment of Motor Imitation (CAMI) using 3D depth cameras. However, incorporating CAMI into serious games and making it accessible in clinic and home settings requires a more scalable approach that uses "off-the-shelf" 2D cameras.
Method: In a brief (one-minute) task, children (23 ASD, 17 typically developing [TD]) imitated a model's dance movements while simultaneously being recorded using Kinect Xbox motion tracking technology (Kinect 3D) and a single 2D camera. Pose-estimation software (OpenPose 2D) was used on the 2D camera video to fit a skeleton to the imitating child. Motor imitation scores computed from the fully automated OpenPose 2D CAMI method were compared to scores computed from the Kinect 3D CAMI and Human Observation Coding (HOC) methods.
Results: Motor imitation scores obtained from the OpenPose 2D CAMI method were significantly correlated with scores obtained from the Kinect 3D CAMI method ($r_{40} = 0.82$, $p < 0.001$) and the HOC method ($r_{40} = 0.80$, $p < 0.001$). Both 2D and 3D CAMI methods showed better discriminative ability than the HOC, with the Kinect 3D CAMI method outperforming the OpenPose 2D CAMI method (area under ROC curve (AUC): $AUC_{HOC} = 0.799$, $AUC_{2D\text{-}CAMI} = 0.876$, $AUC_{3D\text{-}CAMI} = 0.94$). Finally, all motor imitation scores were significantly associated with the social-communication impairment (all $p \leq 0.003$).
Conclusions: This pilot-study demonstrated that motor imitation can be automatically quantified using a single 2D camera.

* Corresponding author at: Center for Neurodevelopmental and Imaging Research, Kennedy Krieger Institute, Baltimore, MD, 21205, USA.
E-mail address: lidstone@kennedykrieger.org (D.E. Lidstone).

https://doi.org/10.1016/j.rasd.2021.101840
Received 4 February 2021; Received in revised form 10 July 2021; Accepted 4 August 2021

Available online 14 August 2021

1. Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder with a core behavioral phenotype defined by severe difficulties in social-communication and excessive repetitive and stereotyped patterns of behavior (American Psychiatric Association, 2013). A meaningful subgroup of children with ASD (~80−90 %) also show more pronounced anomalous development of age-appropriate motor skills (Ament et al., 2015; Bhat & Narayan Bhat, 2020; Craig et al., 2018; Green et al., 2009), the severity of which are reliably associated with the severity of ASD symptoms (Craig et al., 2018; Hirata et al., 2014). Understanding the specificity of motor difficulties in children with ASD, particularly those children with ASD showing disrupted motor development, may provide new therapeutic approaches to improve both social-communicative and motor skills concurrently. Considering autism-associated social and motor difficulties persist throughout the lifespan and negatively affect adaptive skills (Travers et al., 2018), targeted therapeutics in childhood may substantially improve the quality of life of children diagnosed with ASD.

To understand the specificity of motor difficulties in children with ASD, cross-syndrome study designs have been employed revealing sensorimotor difficulties specific to children with ASD. Collectively, cross-syndrome studies show children with ASD are particularly impaired on tasks requiring efficient visual-motor integration (VMI) (Ament et al., 2015; Craig et al., 2018; Dewey, Cantell, & Crawford, 2007; Izawa et al., 2012; Lidstone, Miah, Poston, Beasley, & Dufek, 2020; Lidstone, Miah, Poston, Beasley, Mostofsky et al., 2020; MacNeil & Mostofsky, 2012) with the degree of VMI impairment being significantly associated with more severe impairment in social-communicative skills (Craig et al., 2018; Izawa et al., 2012). For example, on the Movement Assessment Battery for Children (MABC), children with ASD show significant ball catching difficulties, a task that requires rapid integration of visual information to guide motor commands, as compared to typically developing (TD) controls, children with Attention-Deficit Hyperactivity Disorder (ADHD), and children with intellectual disability (Ament et al., 2015; Craig et al., 2018). Further, when learning a novel motor (reaching) task, children with ASD tend to discount visual feedback to a greater extent than children with ADHD and TD controls (Izawa et al., 2012). The underlying nature of the VMI impairment in children with ASD may result from the dynamics of the visual feedback or stimuli. Supporting this hypothesis, Lidstone, Miah, Poston, Beasley, Mostofsky et al. (2020) showed children with ASD are particularly more impaired than children with ADHD, Fetal-Alcohol Spectrum Disorder, and TD controls at adjusting grip-force to track an oscillating dynamic visual target as compared to a stationary visual target (Lidstone, Miah, Poston, Beasley, Mostofsky et al., 2020). Particular difficulty integrating dynamic, but not static, visual information with the motor system is supported by a meta-analysis showing no imitation difficulty in individuals with ASD to emulate static-end point postures, but significant difficulty emulating the dynamic form of the movement (Edwards, 2014). ASD-specific difficulties integrating dynamic visual information to guide motor commands may help explain findings of ASD-specific motor imitation difficulties (Dewey et al., 2007; MacNeil & Mostofsky, 2012). Dewey et al. (2007), showed that while children with ASD, developmental coordination disorder (DCD), and DCD with co-morbid ADHD (DCD + ADHD) all show impaired motor coordination, only children with ASD showed a particular impairment with motor imitation. Similar findings were observed in MacNeil and Mostofsky (2012) where although children with ASD and ADHD showed anomalous basic motor control, children with ASD showed significantly worse motor imitation as compared to both children with ADHD and TD children. Finally, findings suggest that reducing the speed of biological motion during motor imitation reduces the degree of impairment in children with ASD (Lainé, Rauzy, Tardif, & Gepner, 2011), suggesting that the dynamic VMI nature of motor imitation may help explain findings of apparent diagnosis-specific difficulties with dynamic VMI tasks and motor imitation.

Motor imitation is highly desirable as a therapeutic target considering it is a skill crucial for the development of social skills, forming social bonds, and observational learning (Browder, Schoen, & Lentz, 1986; Heyes, 2013; Prinz, 2002). Incorporating motor imitation into serious games may provide a therapeutic approach for concurrently improving social and motor skills in children with ASD. Further, as a promising biomarker of ASD (Tunçgenç et al., 2020), quantifying motor imitation abilities in children with ASD may provide quick and objective tracking of clinically relevant improvements in ASD symptoms from other therapeutic interventions including, but not limited to, behavioral, pharmacological, and non-invasive neuromodulation techniques.

Limiting the rapid and objective quantification of motor imitation is the current reliance on human-observation coding (HOC) that requires time consuming video analysis by trained observers and is therefore slow and quasi-subjective (Mostofsky et al., 2006). Moving towards a more rapid and objective assessment of motor imitation, we recently developed and piloted a computerized assessment of motor imitation (CAMI) (Tunçgenç et al., 2020) designed to quantify the degree of spatial and temporal similarity of joint kinematic trajectories between a child (follower) and a model performer (leader). Crucially, CAMI scores were found to be highly correlated with the HOC method and outperformed HOC at discriminating children with ASD from TD controls (Tunçgenç et al., 2020). The development of CAMI has made an important advancement to the field as an automated, objective, scalable, and valid method of quantifying motor imitation. Scalability of CAMI is currently somewhat limited by the fact that it was validated using three-dimensional kinematic joint trajectories that were obtained using specialized depth-sensing cameras and considerable user-input was required to supervise the semi-automated software skeleton-fitting process (Tunçgenç et al., 2020). The requirement of specialized cameras and user supervised data processing limit the scalability at the levels of assessment and analysis. Promising approaches to increase the scalability and speed of motor imitation assessments are to use a single 2D camera (e.g., webcam) to record motor imitation performance and to perform fully automated pose-estimation on the two-dimensional videos, respectively.

OpenPose is an open-source pose-estimation software (Cao, Simon, Wei, & Sheikh, 2017) that has been reliably used to quantify pathological kinematics from clinical populations using a single 2D camera (OpenPose 2D) (Kidziński et al., 2020). Therefore, automated pose-estimation from 2D video using OpenPose 2D is a promising method to rapidly quantify CAMI. However, the validity of the proposed OpenPose 2D CAMI method needs to be established against the previously validated three-dimensional method (Kinect 3D CAMI) (Tunçgenç et al., 2020).

The purpose of the current pilot study is to examine the construct and concurrent validity of the proposed OpenPose 2D CAMI method by comparing its performance with a "traditional" Human Observation Coding (HOC) approach and a previously validated Kinect 3D CAMI method at quantifying motor imitation in children with ASD and TD controls. We hypothesize that: (a) OpenPose 2D CAMI scores will be highly correlated with Kinect-3D CAMI scores, (b) OpenPose 2D CAMI scores will outperform HOC in diagnosis-discriminative ability; however it will underperform Kinect 3D given the higher spatial dimensionality of the 3D method, and (c) OpenPose 2D CAMI scores will be significantly associated with core ASD social-communicative difficulties.

2. Methods

2.1. Participants

The Johns Hopkins Medical Institutional Review Board approved this experiment. Oral assent and written informed consent were obtained from all participants and their legal guardians.

Participants included 46 children aged 8–12 years (28 ASD, 18 TD). Autism diagnosis was based on DSM-5 criteria and was confirmed on site by research-reliable assessors using the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2), the Autism Diagnostic Interview-Revised (ADI-R). Parent-report of Social Responsiveness Scale (SRS-2) and Repetitive Behavior Scale-Revised (RBS-R) were also obtained. To be included in the study, children needed a Full-Scale IQ score ≥ 80 or at least one index score ≥ 80 (Verbal Comprehension, Visual, Spatial or Fluid Reasoning Index) on the Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V).

Five children were excluded because of technical issues (i.e. limbs of the child moved out of camera view for extended period) and one child was excluded as an outlier in the TD group. Therefore, the final dataset included 40 children (23 ASD, 17 TD) (Table 1).

2.2. Procedures

2.2.1. Dance imitation task

Children completed a 60-second dance imitation task where they were instructed to perform big, whole-body movements to match the movements of a model presented on a large monitor. The children stood on a floor marking that was 1.88 m from the monitor. Two Kinect Xbox depth sensors were positioned in front of and behind the participants and recorded depth data and video concurrently at 30 Hz. Depth data and video captured by the 2D camera recorded simultaneously by the Kinect devices were subsequently used to obtain 3D and 2D kinematics, respectively. Prior to recording data, children completed a familiarization trial to ensure that they understood the task instructions.

2.2.2. OpenPose 2D CAMI

A single digital camera recorded the frontal plane of the participant at 30 Hz (Microsoft Kinect 2.0; 540 × 960 resolution; RGB files; 24 bits per pixel). The video file was spliced with the start defined by the frame where the child initiates their first movement in response to the performer and the end frame defined by the frame where the video presentation ceased. Video recordings (~60 s) were processed locally using freely available OpenPose software demo without modification (https://github.com/CMU-Perceptual-Computing-Lab/openpose) on a laptop (MSI GP63 Leopard 8RE) with a 6GB graphics card (GeForce® GTX 1060). The child was the only person in the video frame with the experimenters hidden behind a barrier during the imitation task. For each video frame, the BODY_25 model in OpenPose tracked 25 key points: nose, neck, mid-hip and bilateral eyes, ears, shoulders, elbows, wrists, hips, knees, ankles, heels, big and small toes. OpenPose tracking and key point export for each 60-second video file was completed in approximately 200 s. Key points for each video frame were output in JSON files that were subsequently used in custom MATLAB scripts for data reduction and post-processing.

In MATLAB, data was reduced from 25 to 15 key points to produce a skeleton that was most like that generated by iPi Mocap, the software used to process the 3D data (Fig. 1). Gaps were automatically filled using linear interpolation for gaps up to a maximum of 30 frames (1 s of data) and key point trajectories were smoothed using a zero-lag 4th order low-pass Butterworth filter with a 5 Hz cut-off (Stenum, Rossi, & Roemmich, 2021) (Fig. 2).

Detailed information regarding key point tracking gaps can be found in the Supplementary material file (Tables S1 and S2 – Supplementary file). Briefly, the maximum number of consecutive gaps were quantified and averaged across participants for each key point. The findings revealed that the mean number of consecutive gaps did not exceed 4 frames for any key point (Table S2 –

Table 1
Participant characteristics.

	ASD ($n = 23$)	TD ($n = 17$)	p-value
Age (years)	9.97 ± 1.31	10.6 ± 1.16	0.13
Sex (M/F)	22/1	14/3	–
ADOS-2 Total	16.09 ± 4.37	–	–
SRS-2 Total Raw Score	98.09 ± 19.21	19.54 ± 11.31	< 0.001
RBS-R Total Score	31.42 ± 18.92	1.13 ± 1.84	< 0.001
WISC-V FSIQ	96.86 ± 15.53	109.29 ± 9.86	0.01
WISC-V Visual Spatial Index (VSI)	103.47 ± 15.70	110.11 ± 14.06	0.17

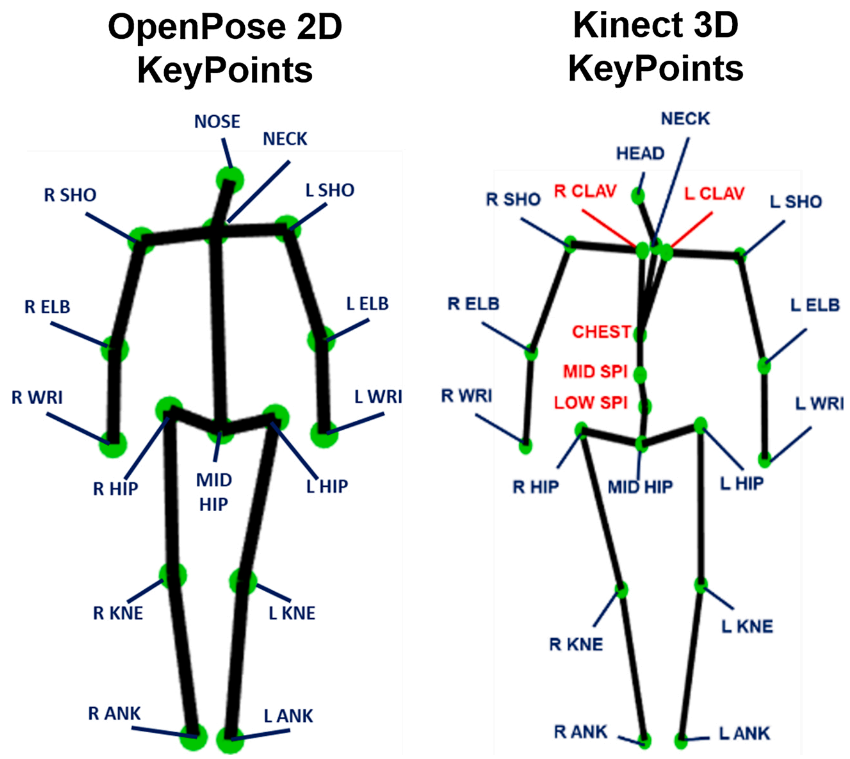

Fig. 1. *Left* - Reduced set ($n = 15$) of OpenPose 2D keypoints from the BODY_25 model. *Right* - iPi 3D Mocap skeleton keypoints ($n = 20$) shown with markers comparable to OpenPose 2D in blue and additional markers unique to Kinect 3D in red (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

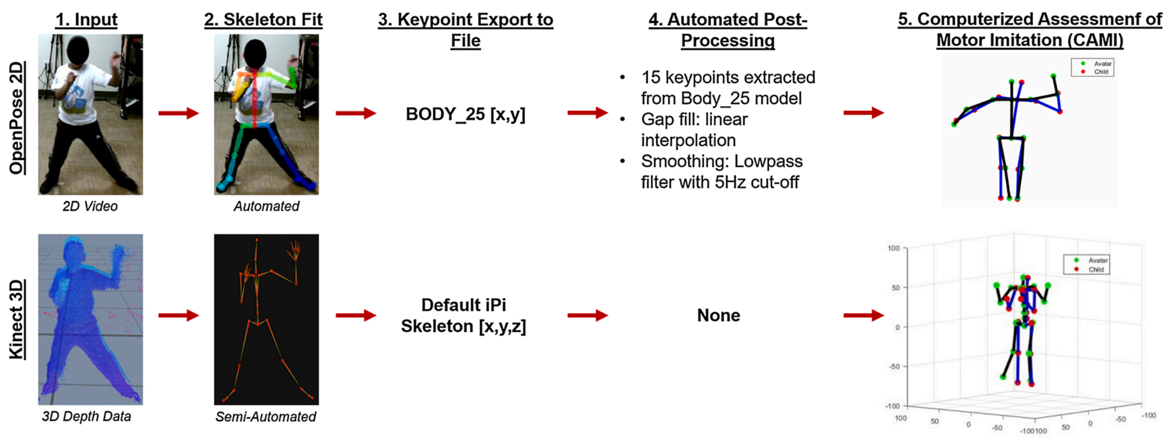

Fig. 2. Summary of OpenPose 2D and Kinect 3D methods from acquisition to Computerized Assessment of Motor Imitation (CAMI).

Supplementary file). In the Supplementary file we further show that the selection of interpolation method (linear vs. previous neighbor interpolation) (Fig. S3 – Supplementary file) had no effect on CAMI scores (Table S3 – Supplementary file).

2.2.3. *Kinect 3D CAMI*

Three-dimensional point cloud data were recorded concurrently with video at 30 Hz using two Microsoft 2.0 Kinect depth sensors positioned in front of and behind the imitating child. The depth files were exported to iPi Recorder software where the two depth files were merged with a calibration file recorded on the day of testing. For calibration, a rectangular board was positioned in view of and mid-way between the front and rear Kinect sensors to estimate camera position and orientation. Merged files were subsequently processed using iPi MoCap Studio software where a skeleton was automatically fit within the point cloud depth data of the imitating child for each frame. However, due to problems with this automated tracking where the skeleton did not align with the depth data, all videos were reviewed by at least two researchers who used a combination of manual editing and the automated tracking feature of the software to ensure the skeleton's perfect fit to the point cloud.

After the depth data was aligned with the skeleton for the entirety of the movement sequence, the motion data was exported using the Biomech add-on of the iPi MoCap Studio software. The x-y-z coordinates of data from the following 20 joints were exported: Hip, LowerSpine, MiddleSpine, Chest, Neck, Head, RClavicle, RShoulder, RForearm, RHand, LClavicle, LShoulder, LForearm, LHand, RThigh, RShin, RFoot, LThigh, LShin and LFoot. No further, post-processing steps were performed on the key point trajectories.

2.2.4. Human observation coding (HOC)

Detailed methods for the traditional HOC method and established inter-rater reliability are provided elsewhere (Tunçgenç et al., 2020). Briefly, trained video-coders scored each movement type ($n = 14$) (Fig. 3) within the dance sequence as correct or incorrect. Within each movement type ($n = 14$) of the dance sequence, correct imitation on the correct side was scored as +1, incorrect imitation was scored as 0, correct movements performed on the reverse side were scored as +0.5, and any unnecessary repetition of a movement resulted in a score of -1, deducted only once per movement type. Scores were summed and divided by the maximum possible score (176 components) to generate a normalized score ranging from 0 (worst imitation) to 1 (perfect imitation).

2.2.5. Computerized Assessment of Motor Imitation (CAMI)

The CAMI algorithm used in this study was previously validated against a human observation coding (HOC) method for quantifying motor imitation in a recent study by our group (Tunçgenç et al., 2020). Here, we provide a brief overview of the CAMI algorithm; detailed overview of the CAMI algorithm is provided elsewhere (Tunçgenç et al., 2020) and the CAMI code can be accessed on GitHub (https://github.com/carolina-pacheco/CAMI).

Three crucial aspects of the CAMI algorithm are: (a) the quantification of both spatial and temporal imitation errors by comparing the joint trajectories of the imitator to the "gold standard" performer; (b) the weighting of joint-specific spatial errors by their relative contribution to each movement sequence completed by the performer; and (c) the linear combination of spatial and temporal errors to generate a single motor imitation score from 0 (worse imitation score) to 1 (best imitation score). The CAMI steps outlined below were the same for OpenPose 2D and Kinect 3D data.

1 **Pre-processing:** The child's and the gold standard's motion data are translated by locating their hips' positions at the origin and the child's limb lengths are normalized to the gold standard's skeleton. In the case of the Kinect 3D data, an additional alignment step was performed where the shoulder keypoint positions from the first frame of the child were rotated to align with the model performer.
2 **Automatic joint importance estimation:** Using the gold standard data, the relative contributions of each joint for each movement type ($n = 14$) are computed based on the amount of displacement observed (Fig. 3). Joints that were displaced more in the gold standard data for a given movement type are considered to contribute more to the movement (e.g., wrist) and hence affected the imitation score more than joints that stayed static (e.g., spine).
3 **Computation of the distance feature**: Using dynamic time warping (DTW) (Sakoe, 1978), the child's time-course was aligned to the model's time-course for the entirety of the sequence by finding a time warp that minimizes the Euclidean distance between them. The DTW distances of each movement type were calculated considering the relative importance of each joint as computed in step-2. The distances for the movement types were then averaged to make up the child's total DTW distance (dist), which was subsequently transformed into a distance score (s_{dist}).
4 **Computation of the time features**: Using the DTW warping path information, time asynchrony features were computed for the entire sequence (Folgado et al., 2018): the duration that children were delayed with respect to the model (t_{delay}) and the duration that children performed the movements in advance of the model (t_{adv}).
5 **Computation of the CAMI score**: Using metric learning techniques, the three variables (s_{dist}, t_{delay}, t_{adv}) were linearly combined to make up the child's imitation score. The weights used for this linear combination were previously determined using Kinect 3D

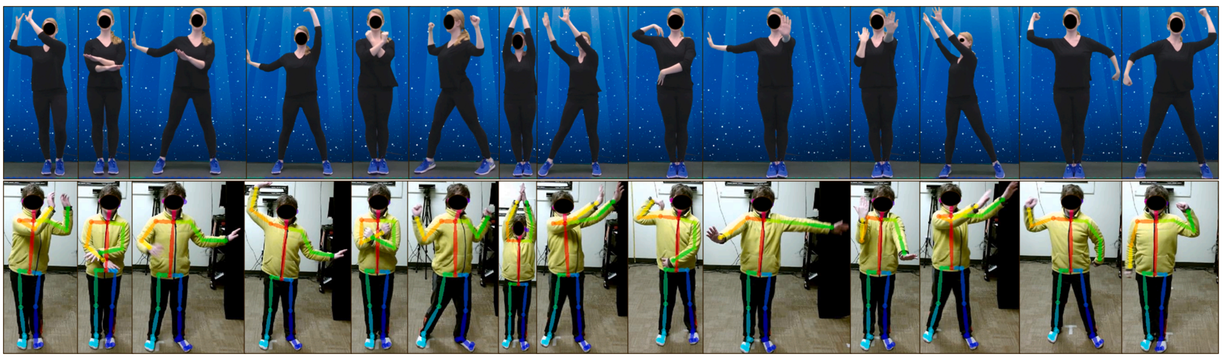

Fig. 3. A single frame from the stimulus video of the model (top) and imitating child (below) shown from each of the 14 movement sequences. The bottom images additionally show the OpenPose BODY_25 key point tracking.

CAMI scores in a data-driven manner using 3-fold cross-validation technique to maximize the correlation between CAMI and HOC (Tunçgenç et al., 2020).

2.3. Statistical procedures

Group differences (ASD vs. TD) in motor imitation scores for HOC, Kinect 3D CAMI, and OpenPose 2D CAMI methods were examined using independent sample t-tests ($\alpha = 0.05$). The construct (OpenPose 2D CAMI vs. HOC) concurrent validity (OpenPose 2D CAMI vs. Kinect 3D CAMI) of the OpenPose 2D CAMI method were examined via correlation analysis.

2.3.1. Diagnostic classification ability

To examine the diagnosis discriminative ability of all three motor imitation scoring methods (HOC, Kinect 3D CAMI, and OpenPose 2D CAMI), we trained a machine learning algorithm (linear support vector machines [SVMs] using 3-fold cross-validation) to classify subjects into their diagnostic groups (ASD/TD) using their imitation scores-only (Tunçgenç et al., 2020). The k-fold cross-validation procedure was used to prevent overfitting of the classifier. Further, we generated receiver-operating characteristic (ROC) curves for each imitation scoring method for each imitation scoring method to study the relationship between true positive and false positive rates at different classification regimes. The area under the ROC curve (AUC) was calculated to measure the discriminative ability of the SVM with a larger AUC indicating better discriminative ability.

3. Results

3.1. Concurrent and construct validity of OpenPose 2D CAMI

OpenPose 2D CAMI scores and Kinect 3D CAMI scores were significantly correlated demonstrating concurrent validity of the OpenPose 2D method. The OpenPose 2D method also showed construct validity showing strong correspondence with HOC motor imitation scores (OpenPose 2D CAMI vs. Kinect 3D CAMI: $r_{40} = 0.82$, $p < 0.001$; OpenPose 2D CAMI vs. HOC: $r_{40} = 0.80$, $p < 0.001$; Kinect 3D CAMI vs. HOC: $r_{40} = 0.69$, $p < 0.001$) (Fig. 4A–C).

3.2. Diagnostic discriminative ability of CAMI vs. HOC

Whereas all three motor imitation scoring methods showed worse imitation performance in children with ASD as compared to TD controls (ASD < TD, all $p \leq 0.001$; Fig. 5A; Table 2), the Kinect 3D CAMI method was superior to all methods in discriminative ability (ROC AUC: Kinect 3D > OpenPose 2D CAMI > HOC) that was supported by higher classification accuracy of Kinect 3D CAMI as compared to all methods with OpenPose 2D CAMI outperforming the traditional HOC method ($\text{accuracy}_{\text{HOC}} = 75$ %, $\text{accuracy}_{\text{CAMI-2D}} = 77.5$ %, $\text{accuracy}_{\text{CAMI-3D}} = 85$ %; Kinect 3D CAMI > OpenPose 2D CAMI > HOC) (Fig. 5B).

Mean accuracy and 95 % confidence intervals (CI) were also provided by repeating the 3-fold cross-validation 100 iterations: (a) randomly dividing data into 3 folds, (b) running 3-fold cross-validation, and (c) reporting average performance of the classifier. Results were similar with mean classification accuracy highest for the Kinect 3D CAMI method followed by the OpenPose 2D CAMI and HOC methods (mean $\text{accuracy}_{\text{HOC}} = 68.2$ % [95 % CI: 53.7–73.2 %], mean $\text{accuracy}_{\text{CAMI-2D}} = 76.5$ % [95 % CI: 70.7–80.5 %], mean

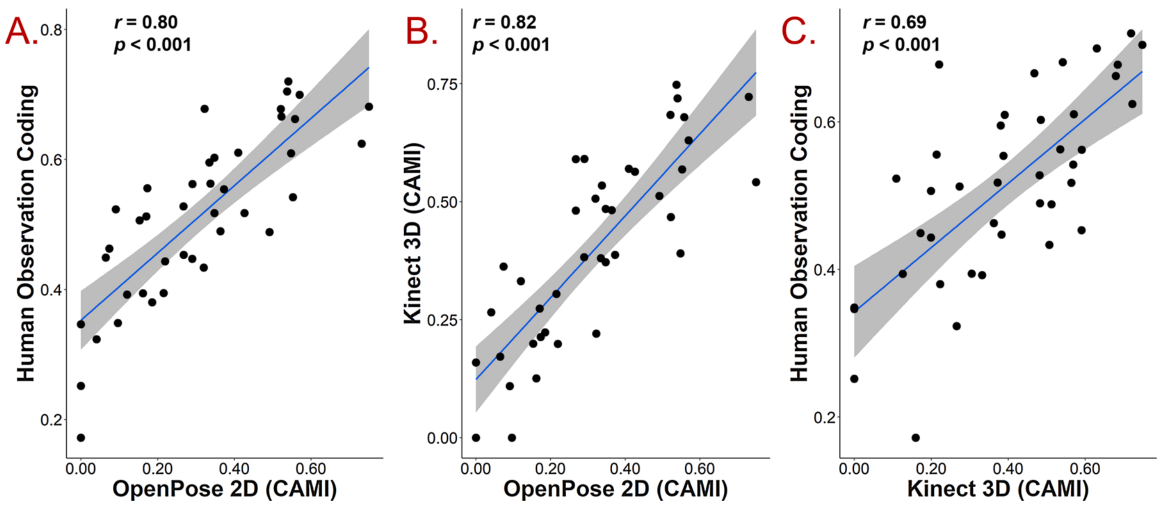

Fig. 4. **(A)** Correlation between Human Observation Coding (HOC) scores and OpenPose 2D CAMI scores ($r_{40} = 0.80$, $p < 0.001$). **(B)** Correlation between OpenPose 2D CAMI scores and Kinect 3D CAMI scores ($r_{40} = 0.82$, $p < 0.001$). **(C)** Human Observation Coding (HOC) scores and Kinect 3D CAMI scores ($r_{40} = 0.69$, $p < 0.001$).

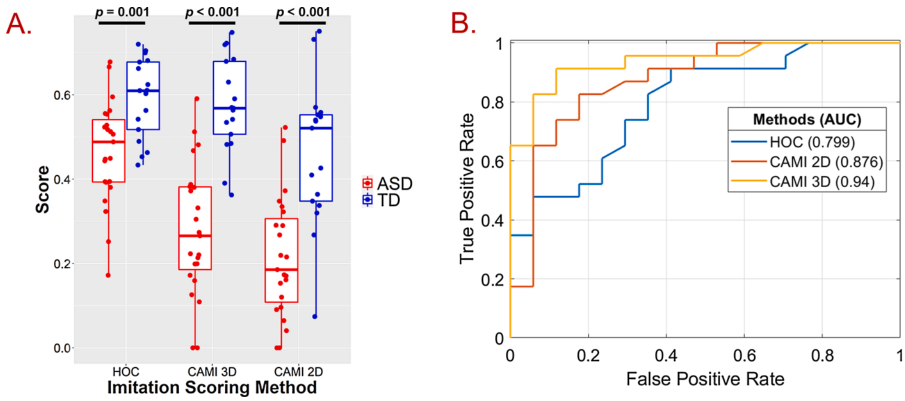

Fig. 5. **(A)** Group differences in motor imitation scores derived from Human Observation Coding (HOC) (TD > ASD: $p = 0.001$, Kinect 3D CAMI (TD>ASD: $p < 0.001$), and OpenPose 2D CAMI scores (TD>ASD: $p < 0.001$). **(B)** Receiver-operating characteristic curves: true positive rate vs. false positive rate as classification threshold is varied. The area under the curve (AUC) indicates the diagnostic ability for each of the 3 motor imitation methods (best possible AUC is 1, meaning 0 false positives and 100 % true positives).

Table 2
Mean and standard deviation of motor imitation scores obtained using Human Observation Coding (HOC), Kinect 3D CAMI, and OpenPose 2D CAMI methods for children with ASD and TD controls.

	ASD ($n = 23$)	TD ($n = 17$)	p-value
Human Observation Coding	0.46 ± 0.12	0.59 ± 0.09	0.001
OpenPose 2D (CAMI)	0.21 ± 0.14	0.46 ± 0.16	< 0.001
Kinect 3D (CAMI)	0.27 ± 0.15	0.57 ± 0.11	<0.001

$\text{accuracy}_{\text{CAMI-3D}} = 82.3$ % [95 % CI: 75.6–85.4 %]; Kinect 3D CAMI > OpenPose 2D CAMI > HOC).

3.3. Motor imitation correlations with core autism symptoms and performance IQ

Social-communicative impairment (SRS-2 total raw scores) were significantly associated with motor imitation from all three methods with Kinect 3D CAMI showing the strongest associations with SRS-2 and OpenPose 2D CAMI outperforming HOC (Kinect 3D CAMI vs. SRS-2: $r_{35} = -0.66$, $p < 0.001$; OpenPose 2D CAMI vs. SRS-2: $r_{35} = -0.57$, $p = 0.002$; HOC vs. SRS-2: $r_{35} = -0.49$, $p = 0.003$) (Fig. 6A–C). Repetitive Behavior Scale-Revised (RBS-R) total scores were also significantly associated with motor imitation scores (Kinect 3D CAMI vs. RBS-R: $r_{36} = -0.62$, $p < 0.001$; OpenPose 2D vs. RBS-R: $r_{36} = -0.62$, $p < 0.001$; HOC vs. RBS-R: $r_{36} = -0.52$, $p =$

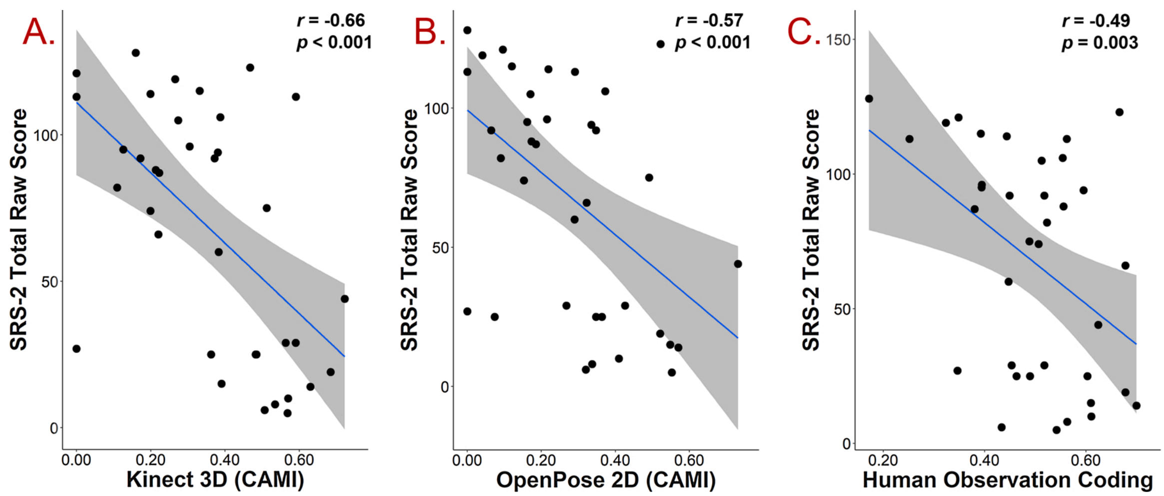

Fig. 6. **(A)** Correlation between Kinect 3D CAMI scores and SRS-2 scores ($r_{35} = -0.66$, $p < 0.001$). **(B)** Correlation between OpenPose 2D CAMI scores and SRS-2 scores ($r_{35} = -0.57$, $p < 0.001$). **(C)** Correlation between Human Observation Coding (HOC) scores and SRS-2 scores ($r_{35} = -0.49$, $p = 0.003$).

0.001) (Fig. S1 – Supplementary file). Finally, motor imitation scores showed no significant associations with the WISC-V Visual Spatial Index (VSI) (Kinect 3D CAMI vs. VSI: $r_{40} = 0.17$, $p = 0.30$; OpenPose 2D vs. VSI: $r_{40} = -0.03$, $p = 0.84$; HOC vs. VSI: $r_{40} = 0.06$, $p = 0.68$) (Fig. S2 – Supplementary file).

4. Discussion

The purpose of the current pilot study was to examine the construct and concurrent validity of using a 2D video and freely available pose-estimation software to rapidly quantify motor imitation as compared to a "traditional" HOC and previously validated Kinect 3D CAMI method (Tunçgenç et al., 2020). In support of the stated hypotheses, CAMI scores derived from the OpenPose 2D method: (a) were significantly correlated with CAMI scores derived from the Kinect 3D method, (b) showed improvement over HOC in terms of diagnosis-discriminative performance with worse performance as compared to the Kinect 3D CAMI method, and (c) were significantly associated with social-communicative skills similar to that observed for both HOC and Kinect 3D CAMI scores. The OpenPose 2D CAMI method requires no user input offering a major advantage over HOC and the Kinect 3D CAMI method in terms of potential for therapeutic utility via incorporation of CAMI into serious games.

The current pilot study provides three key advancements to the assessment of motor imitation: (i) processing of the video data has been fully automated, (ii) use of a single, 2D camera that is low-cost and highly accessible, and (iii) the CAMI scores are obtained speedily within a few minutes. These advancements, particularly the use of a single 2D camera, enhance the utility of the CAMI method to assess motor imitation performance as well as for providing immediate feedback of performance to the clinician, care giver and/or child for use as an intervention tool. Moving forward, it seems entirely feasible that a serious game could be developed to: (a) present a video of a person or avatar performing a series of movement sequences on a large monitor, (b) record the imitating child using a single 2D camera including webcams embedded in computers/mobile devices, (c) perform pose estimation of the 2D recordings using OpenPose or similar systems, and (d) use CAMI to quantify motor imitation – all in the order of a few minutes. The flexibility of this scalable approach to assessing motor imitation may be particularly valuable to clinicians serving children with ASD in rural communities where access to care is significantly more challenging as compared to metropolitan communities (Antezana, Scarpa, Valdespino, Albright, & Richey, 2017). However, moving towards the therapeutic utility of CAMI, it remains crucial to examine longitudinally whether CAMI sufficiently captures changes in motor imitation over time and whether improvements in motor imitation correspond with improvements in social-communicative skills in children with ASD.

Whereas our proposed approach for improving imitation skills uses videos of pre-recorded dance sequences and "off-line" computation of motor imitation performance using CAMI, several groups have examined the use of robots to perform a small number of gestures, record and classify gestures from the imitating child, and adapt movements or instruction to the child based on whether the child correctly performs the gesture (Nguyen et al., 2020; Zheng et al., 2016). In these approaches, the robot performs gestures while pose estimation algorithms coupled with machine learning algorithms classify the gestures performed by the child to make subsequent decisions to promote learning (e.g., provide instruction > wait for child to perform the correct gesture > modify the movement). This approach to improving imitation using adaptable gameplay and robot interaction is intriguing but comes with some major limitations. First, the most obvious limitation of robot-mediated imitation are the costs and mechanical constraints (e.g., speed, coordination). Motor imitation of a human model performer overcomes the obvious mechanical limitations of robot-mediated imitation and is more generalizable to actual social situations with other humans. Second, gesture recognition during robot-mediated imitation is based on the gesture end-points or very simple gesture sequences that can be easily classified using machine learning algorithms. Evidence suggests that dynamic motor imitation as compared to static, or end-point, imitation is most impaired in children with ASD (Edwards, 2014). Therefore, examining the imitation of dynamic movement sequences, particularly sequences requiring simultaneous movement with multiple effectors (McAuliffe, Pillai, Tiedemann, Mostofsky, & Ewen, 2017), may have greater clinical relevance than static "end-point" imitation. The methods described in the current study for both recording and analyzing motor imitation using 2D cameras overcomes the limitations of robot-mediated imitation by using videos of human models and computing the motor imitation of complex movement sequences using automated methods. The use of videos of human models is also advantageous over face-to-face imitation of a human model to promote consistency and standardization between participants (inter-subject reliability) and across consecutive clinical assessments (inter-session reliability).

Our proposed approach of using video movement sequences and calculating motor imitation "off-line" using CAMI is more scalable and flexible than what can be provided from human-robot interaction approaches. A future direction will be to incorporate the proposed OpenPose 2D CAMI method into serious games for rapid quantification of motor imitation and adaptation of presented visual stimuli to guide improvements in motor imitation skill. Considering that our approach for quantifying motor imitation requires time-series measures of the spatial and temporal similarity with the "gold standard" video, incorporation of adaptation and instruction would have to be provided after a series of trials resulting from the need to process CAMI scores following completion of the sequence. Future studies are needed to examine the therapeutic utility of a serious game that automatically (a) generates CAMI scores from 2D video and (b) guides learning using a feedback-driven implicit and/or explicit instruction delivery method to progressively increase motor imitation skills in children with ASD. This proposed flexible approach to developing motor imitation-focused games would allow clinicians to create their own video stimuli of specific social and motor skills for which the child shows difficulties and guide skill learning in an automatic stepwise feedback-driven manner. Manipulating the speed of the visual stimuli (Lainé et al., 2011; Tardif, Lainé, Rodriguez, & Gepner, 2007) and/or the complexity of the movement sequences (McAuliffe et al., 2017) are two proposed methods to adjust the difficulty of motor imitation to promote skill acquisition.

In addition to incorporating OpenPose 2D CAMI into automated and adaptive serious games to promote learning of motor imitation skills, future studies should develop methods to improve the diagnostic discriminability of the 2D CAMI method. Our findings show that the diagnostic discriminative ability of Kinect 3D CAMI scores was superior to the OpenPose 2D CAMI scores. There are a few possible reasons for this outcome. First, the spatial dimensionality of the data is diminished using 2D vs. 3D kinematics and this deficit may contribute to the loss of diagnostic-discriminative performance. Second, in contrast to the Kinect 3D method, no manual pre-processing of tracked OpenPose 2D kinematics were performed. While a fully automated pose-estimation pipeline is crucial for clinical and therapeutic utility of CAMI, the process will undoubtedly result in increased tracking errors. Future studies should examine new methods to improve upon OpenPose 2D pose-estimation algorithm to mitigate tracking errors. Elucidating the factors contributing to the poorer discriminative ability of the 2D CAMI method as compared to the 3D CAMI method are important to optimize our proposed 2D CAMI approach and its clinical and therapeutic utility.

Finally, while other serious game machine learning digital phenotyping methods have shown efficacy for accurately classifying autism in children using features such as tablet hand gestures (Anzulewicz, Sobota, & Delafield-Butt, 2016) and reaching kinematics (Cavallo et al., 2021), such features have not been shown to be particularly disrupted in children with ASD nor associated with core ASD symptoms. In contrast, motor imitation has been shown to be particularly disrupted in children with ASD as compared to TD controls and children with ADHD (MacNeil & Mostofsky, 2012) and CAMI is significantly associated with core ASD symptoms (Tunçgenç et al., 2020). CAMI therefore objectively quantifies a feature particularly disrupted in children with ASD, captures ASD symptom severity, and classifies ASD diagnosis with high-accuracy – all key pieces moving towards an endophenotype that can be used to improve diagnosis of ASD and quantitative tracking of the effects of behavioral and/or pharmaceutical intervention.

4.1. Limitations

This pilot study has some limitations. First, the relatively small sample size and the examination of a single imitation trial limits the generalizability of our findings. However, despite our low sample size, our classifier was generated using a low feature-to-sample ratio (1/26; 1 feature/26 training samples) and therefore at low risk for classifier overfitting using k-fold cross validation (Vabalas, Gowen, Poliakoff, & Casson, 2019). Second, the test-retest reliability of OpenPose 2D CAMI scores were not explored in this study. Third, this initial study did not include longitudinal data which would provide for examining whether the OpenPose 2D CAMI is sensitive to changes in motor imitation skill over time. The flexibility of the OpenPose 2D method to quantify imitation from videos acquired at home may improve our ability to track changes in motor imitation longitudinally without the need for frequent laboratory visits. Finally, eye-tracking was not included as part of the motor imitation assessment that could provide an objective measure of task engagement.

5. Conclusion

In conclusion, this pilot study shows that motor imitation can be automatically quantified, in the order of a few minutes, using a single 2D camera and fully automated pose estimation software (OpenPose 2D CAMI) with improved performance over motor imitation scored by an HOC method. Further, we show that CAMI measured using the OpenPose 2D CAMI method is able to discriminate a group of children with ASD from TD controls, with the degree of motor imitation impairment being predictive of poorer social-communication skills (i.e., increased core ASD severity). The proposed approach of using 2D pose-estimation for automated processing of motor imitation is highly flexible and accessible for clinicians with minimal technical experience. This study thereby provides a methodological framework for designing a serious game to capture longitudinal changes in motor imitation skills using CAMI at home and in clinical settings. The pilot data presented here provides a foundation for employing these methods in a therapeutic context to enhance brain systems crucial to improving imitative skills that are central to the development of a range of motor, adaptive and social-communicative abilities.

CRediT authorship contribution statement

Daniel E. Lidstone: Wrote the manuscript, Carried out the statistical analysis. **Rebecca Rochowiak:** Collected and pre-processed the data. Designed the experiments and the CAMI method. **Carolina Pacheco:** Wrote the manuscript. Carried out the statistical analysis. Designed the experiments and the CAMI method. **Bahar Tunçgenç:** Wrote the manuscript. Collected and pre-processed the data. Designed the experiments and the CAMI method. **Rene Vidal:** Designed the experiments and the CAMI method. **Stewart H. Mostofsky:** Wrote the manuscript. Designed the experiments and the CAMI method. All authors contributed intellectually to this work.

Funding

Recruitment for this study was supported by National Institutes of Health (NIH) R01 grant MH106564-02 and Specialized Center grant P50HD103538, while algorithm development was supported by NIH R01 grant HD87133-01 and the Simons Foundation for Autism Research Initiative (SFARI).

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

The authors would like to thank all the children and their families for participating in this study. The authors would also like to acknowledge Dr. Jan Stenum for assisting with the implementation of OpenPose software with our dataset.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.rasd.2021.101840.

References

Ament, K., Mejia, A., Buhlman, R., Erklin, S., Caffo, B., Mostofsky, S., & Wodka, E. (2015). Evidence for specificity of motor impairments in catching and balance in children with autism. *Journal of Autism and Developmental Disorders, 45*(3), 742–751. https://doi.org/10.1007/s10803-014-2229-0.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Antezana, L., Scarpa, A., Valdespino, A., Albright, J., & Richey, J. A. (2017). Rural trends in diagnosis and services for autism spectrum disorder. *Frontiers in Psychology*, (April 20)https://doi.org/10.3389/fpsyg.2017.00590. Frontiers Research Foundation.

Anzulewicz, A., Sobota, K., & Delafield-Butt, J. T. (2016). Toward the Autism Motor Signature: Gesture patterns during smart tablet gameplay identify children with autism. *Scientific Reports, 6*. https://doi.org/10.1038/srep31107.

Bhat, A. N., & Narayan Bhat, A. (2020). *Is motor impairment in autism Spectrum disorder distinct from developmental coordination disorder? A report from the SPARK study*. Retrieved from https://academic.oup.com/ptj/article-abstract/100/4/633/5801997.

Browder, D., Schoen, S. F., & Lentz, F. E. (1986). Learning to learn through observation. *Journal of Abnormal Child Psychology, 20*(4), 447–461.

Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2D pose estimation using part affinity fields. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 1302–1310. https://doi.org/10.1109/CVPR.2017.143, 2017-Janua (Xxx).

Cavallo, A., Romeo, L., Ansuini, C., Battaglia, F., Nobili, L., Pontil, M., … Becchio, C. (2021). Identifying the signature of prospective motor control in children with autism. *Scientific Reports, 11*(1), 3165. https://doi.org/10.1038/s41598-021-82374-2.

Craig, F., Lorenzo, A., Lucarelli, E., Russo, L., Fanizza, I., & Trabacca, A. (2018). Motor competency and social communication skills in preschool children with autism spectrum disorder. *Autism Research, 11*(6), 893–902. https://doi.org/10.1002/aur.1939.

Dewey, D., Cantell, M., & Crawford, S. G. (2007). Motor and gestural performance in children with autism spectrum disorders, developmental coordination disorder, and/or attention deficit hyperactivity disorder. *Journal of the International Neuropsychological Society, 13*(2), 246–256. https://doi.org/10.1017/S1355617707070270.

Edwards, L. A. (2014). A meta-analysis of imitation abilities in individuals with autism spectrum disorders. *Autism Research, 7*(3), 363–380. https://doi.org/10.1002/aur.1379.

Folgado, D., Barandas, M., Matias, R., Martins, R., Carvalho, M., & Gamboa, H. (2018). Time alignment measurement for time series. *Pattern Recognition, 81*, 268–279. https://doi.org/10.1016/j.patcog.2018.04.003.

Green, D., Charman, T., Pickles, A., Chandler, S., Loucas, T., Simonoff, E., & Baird, G. (2009). Impairment in movement skills of children with autistic spectrum disorders. *Developmental Medicine and Child Neurology, 51*(4), 311–316. https://doi.org/10.1111/j.1469-8749.2008.03242.x.

Heyes, C. (2013). What can imitation do for cooperation?. *Cooperation and its evolution* (pp. 313–331). Cambridge, MA, US: The MIT Press.

Hirata, S., Okuzumi, H., Kitajima, Y., Hosobuchi, T., Nakai, A., & Kokubun, M. (2014). Relationship between motor skill and social impairment in children with autism spectrum disorders. *International Journal of Developmental Disabilities, 60*(4), 251–256. https://doi.org/10.1179/2047387713Y.0000000033.

Izawa, J., Pekny, S. E., Marko, M. K., Haswell, C. C., Shadmehr, R., & Mostofsky, S. H. (2012). Motor learning relies on integrated sensory inputs in ADHD, but over-selectively on proprioception in autism spectrum conditions. *Autism Research, 5*(2), 124–136. https://doi.org/10.1002/aur.1222.

Kidziński, Ł., Yang, B., Hicks, J. L., Rajagopal, A., Delp, S. L., & Schwartz, M. H. (2020). Deep neural networks enable quantitative movement analysis using single-camera videos. *Nature Communications, 11*(1), 1–10. https://doi.org/10.1038/s41467-020-17807-z.

Lainé, F., Rauzy, S., Tardif, C., & Gepner, B. (2011). Slowing down the presentation of facial and body movements enhances imitation performance in children with severe autism. *Journal of Autism and Developmental Disorders, 41*(8), 983–996. https://doi.org/10.1007/s10803-010-1123-7.

Lidstone, D. E., Miah, F. Z., Poston, B., Beasley, J. F., & Dufek, J. S. (2020). Manual dexterity in children with autism spectrum disorder: A cross-syndrome approach. *Research in Autism Spectrum Disorders, 73*(March), Article 101546. https://doi.org/10.1016/j.rasd.2020.101546.

Lidstone, D. E., Miah, F. Z., Poston, B., Beasley, J. F., Mostofsky, S. H., & Dufek, J. S. (2020). Children with autism spectrum disorder show impairments during dynamic versus static grip-force tracking. *Autism Research, 13*(12), 2177–2189. https://doi.org/10.1002/aur.2370.

MacNeil, L. K., & Mostofsky, S. H. (2012). Specificity of dyspraxia in children with autism. *Neuropsychology, 26*(2), 165–171. https://doi.org/10.1037/a0026955.

McAuliffe, D., Pillai, A. S., Tiedemann, A., Mostofsky, S. H., & Ewen, J. B. (2017). Dyspraxia in ASD: Impaired coordination of movement elements. *Autism Research, 10*(4), 648–652. https://doi.org/10.1002/aur.1693.

Mostofsky, S. H., Dubey, P., Jerath, V. K., Jansiewicz, E. M., Goldberg, M. C., & Denckla, M. B. (2006). Developmental dyspraxia is not limited to imitation in children with autism spectrum disorders. *Journal of the International Neuropsychological Society, 12*(3), 314–326. https://doi.org/10.1017/S1355617706060437.

Nguyen, S. M., Collot-Lavenne, N., Guillon, S., Tula, P., Paez, A., Bouaida, M., … Lohr, C. (2020). An implementation of an imitation game with ASD children to learn nursery rhymes. *arXiv*. Retrieved from http://acm.org/about/class/1998/]:.

Prinz, W. (2002). Experimental approaches to imitation. *The imitative mind* (pp. 143–162). Cambridge University Press. https://doi.org/10.1017/CBO9780511489969.009.

Sakoe, H. (1978).). Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*.

Stenum, J., Rossi, C., & Roemmich, R. T. (2021). Two-dimensional video-based analysis of human gait using pose estimation. *PLoS Computational Biology, 17*(4). https://doi.org/10.1371/journal.pcbi.1008935.

Tardif, C., Lainé, F., Rodriguez, M., & Gepner, B. (2007). Slowing down presentation of facial movements and vocal sounds enhances facial expression recognition and induces facial-vocal imitation in children with autism. *Journal of Autism and Developmental Disorders, 37*(8), 1469–1484. https://doi.org/10.1007/s10803-006-0223-x.

Travers, B. G., Bigler, E. D., Duffield, T. C., Prigge, M. D. B., Froehlich, A. L., Lange, N., … Lainhart, J. E. (2018). Longitudinal development of manual motor ability in autism spectrum disorder from childhood to mid-adulthood relates to adaptive daily living skills. *Developmental Science, 20*(4). https://doi.org/10.1111/desc.12401.

Tunçgenç, B., Pacheco, C., Rochowiak, R., Nicholas, R., Rengarajan, S., Zou, E., … Mostofsky, S. H. (2020). Computerised Assessment of Motor Imitation (CAMI) as a scalable method for distinguishing children with autism. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. https://doi.org/10.1016/j.bpsc.2020.09.001.

Vabalas, A., Gowen, E., Poliakoff, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PLoS One, 14*(11), 1–20. https://doi.org/10.1371/journal.pone.0224365.

Zheng, Z., Young, E. M., Swanson, A. R., Weitlauf, A. S., Warren, Z. E., & Sarkar, N. (2016). Robot-mediated imitation skill training for children with autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 24*(6), 682–691. https://doi.org/10.1109/TNSRE.2015.2475724.