



HHS Public Access

Author manuscript

Biol Psychiatry Cogn Neurosci Neuroimaging. Author manuscript; available in PMC 2022 March 01.

Published in final edited form as:

Biol Psychiatry Cogn Neurosci Neuroimaging. 2021 March ; 6(3): 321–328. doi:10.1016/j.bpsc.2020.09.001.

Computerised Assessment of Motor Imitation (CAMI) as a scalable method for distinguishing children with autism

Bahar Tunçgenç^{1,2}, Carolina Pacheco^{3,4}, Rebecca Rochowiak¹, Rosemary Nicholas⁵, Sundararaman Rengarajan⁶, Erin Zou¹, Brice Messenger¹, Rene Vidal^{3,4}, Stewart H. Mostofsky^{1,7,8}

¹Center for Neurodevelopment and Imaging Research, Kennedy Krieger Institute. 716 N. Broadway, Baltimore, Maryland, 21205, USA

²School of Psychology, University of Nottingham, University Park, Nottingham, NG7 2RD, UK

³Mathematical Institute for Data Science, Johns Hopkins University, 3400 N Charles St., Baltimore, Maryland, 21218, USA

⁴Department of Biomedical Engineering, Johns Hopkins University, 3400 N Charles St., Baltimore, Maryland, 21218, USA

⁵Sir Peter Mansfield Imaging Centre, University of Nottingham, University Park, Nottingham, NG7 2RD, UK

⁶San Diego Joint Doctoral Program in Language and Communicative Disorders, San Diego State University/ University of California

⁷Department of Neurology, Johns Hopkins University School of Medicine, 600 N Wolfe Str., Meyer 6-113, Baltimore, Maryland, 21287, USA

⁸Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, United States, 733 N Broadway, Baltimore, Maryland, 21205, USA

Abstract

Background—Imitation deficits are prevalent in autism spectrum conditions (ASC) and are associated with core autistic traits. Imitating others' actions is central to the development of social skills in typically-developing (TD) populations, as it facilitates social learning and bond formation. We present a Computerised Assessment of Motor Imitation (CAMI) using a brief (one-minute), highly-engaging videogame task.

Methods—Using Kinect Xbox motion tracking technology, we recorded 48 children (27 ASC, 21 TD) as they imitated a model's dance movements. We implemented an algorithm based on metric

Corresponding author: Bahar Tunçgenç, bahar.tuncgenc@nottingham.ac.uk, +44 7450 502920, School of Psychology, University of Nottingham, University Park, NG7 2RD Nottingham, United Kingdom.

Financial Disclosures

The authors report no biomedical financial interests or potential conflicts of interest.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

learning and dynamic time warping (DTW) that automatically detects and evaluates the important joints and returns a score considering spatial position and timing differences between the child and the model. To establish construct validity and reliability, we compared imitation performance measured by the CAMI method to more traditional human observation coding (HOC) method across repeated trials and two different movement sequences.

Results—Results revealed poorer imitation in ASC than TD children ($p < .005$), with poorer imitation being associated with increased core autism symptoms. While strong correlations between the CAMI and HOC methods ($r = .69-.87$) confirmed CAMI's construct validity, CAMI scores classified the children into diagnostic groups better than the HOC scores (accuracy_{CAMI} = 87.2%, accuracy_{HOC} = 74.4%). Finally, by comparing repeated movement trials, we demonstrated high test-retest reliability of CAMI ($r = .73-.86$).

Conclusions—Findings support CAMI as an objective, highly scalable, directly interpretable method for assessing motor imitation differences, providing a promising biomarker for defining biologically meaningful ASC subtypes and guiding intervention.

Keywords

autism; imitation; machine learning; intervention; motor learning; social behaviour

Introduction

Imitating others' actions is crucial for social bond formation and learning [1,2,3], with atypical imitation indicating social-communicative impairments in autism spectrum conditions (ASC) [4,5,6,7,8]. The current standard in imitation assessment is manual human observation coding (HOC), which is subjective, time-consuming and requires intensive coder training. These drawbacks render HOC impractical for use in clinics and home settings. Automatic assessment of imitation is challenging because human motion data are highly heterogeneous (e.g., range of movements is virtually unlimited), high-dimensional (involves spatial and temporal aspects), and human supervision (e.g., expert knowledge) is limited and error-prone. Addressing these issues, this paper presents a Computerised Assessment of Motor Imitation (CAMI) that can improve diagnosis and treatment efforts by providing an objective, continuous and scalable score of imitation performance.

Examining imitation performance with HOC methods requires identifying individual steps involved in an action, the action's style, order of occurrence, repetitions and end-goal if it exists [9,10,11,12]. Participants receive an ordinal score depending on the correct actions and errors made. Thus, the accuracy and precision (i.e., sampling frequency) of HOC is restricted by human subjectivity. Firstly, what constitutes a 'good enough resemblance' to the target action is at the human observers' discretion. Moreover, the defined action categories may likely miss preliminary forms of that action (e.g., flexing the fingers wide open without moving the hand when assessing the action of waving). Finally, assessments are subjective even within agreed-upon standards, an issue often circumvented by seeking high inter-rater reliability from multiple coders. Although this work-around alleviates subjectivity, it adds to coder training and assessment time. As such, the HOC method is

largely confined to research settings and is not conducive for practical use as a diagnostic or treatment tool.

Prior attempts to develop automated methods for assessing imitation performance have focused on determining the match between a participant's movements and those of a template. Two most commonly-used methods are rule-based algorithms [13,14] and algorithms based on dynamic time warping (DTW) [15,16]. Similar to HOC, rule-based algorithms require the researchers to manually define a set of rules. How well the participants meet these rules is automatically assessed by the algorithm. Despite its demonstrated utility in robot-mediated therapy settings with children with ASC [13,14], rule-based methods have very limited generalisability as they require a priori human input for selecting the rules specific to the gestures under study.

In contrast, DTW-based methods assess the spatial similarity between two time-series after correcting for discrepancies in the temporal dimension [17] without requiring human input. DTW-based approaches have been widely used in gesture recognition tasks, where a decision about which gesture the participant performed is outputted based on the similarity between the participant and the template [18,19,20,21]. Existing DTW-based imitation assessment approaches define a metric that is either dichotomous [15] (imitated vs not) or categorical [16] (good vs bad performance), thereby not utilising the continuous distance metric obtained from DTW. These approaches can only capture relatively large variations in imitation performance due to having categorical outputs, an issue that becomes even more prominent in clinical populations such as ASC that display high behavioural variability. Moreover, given the importance of timing in social coordination [22] and in characterising autism-specific imitation impairments [23,24], a valid imitation assessment system must also consider temporal differences.

Machine learning techniques that learn motor patterns and classify individuals into diagnostic groups is another popular approach [25,26,27]. Yet, prior studies did not directly assess imitation ability even when some tasks involved imitation [25,26]. Therefore, it is unclear if the observed differences represent general motor abnormalities or specific imitation impairments. This lack of specificity restricts the use of these methods for intervention purposes.

Characterising and addressing imitation impairments is crucial because imitation plays an important role in social bond formation and learning [1,2,3], joint attention [12], children's play initiation [28], social affiliation and prosocial behaviours [29,30]. Extant research shows that as compared to their typically-developing (TD) peers, imitation in ASC is less frequent, less precise and more delayed [9,10,31,32]. These imitation deficits are more pronounced when the actions appear meaningless or lack an obvious end-goal [5,6,9,11,24]. Impaired imitation is associated with poorer social-communicative functioning in ASC as demonstrated in social responsiveness, social attention, engagement in joint play and social reciprocity [33,34,35].

An automated method that (a) specifically measures imitation performance, (b) does not require manual feature selection, (c) generalises to a range of movement types, (d) provides

continuous scores, and (e) integrates spatial and timing differences would significantly improve diagnosis efforts and robot-mediated and other social-communication interventions for autism.

The CAMI method that we developed uses 3D motion data obtained from sensorless Kinect Xbox cameras. Using DTW and metric learning techniques, CAMI considers differences in both motion trajectories and timing differences. We applied CAMI to a data set comprised of 48 children (27 ASC, 21 TD) as they imitated the dance-like movements of a video model. In this paper, we report on the construct validity of the CAMI method, assessed by comparing the children's CAMI scores to their imitation scores obtained by the HOC method. We also established the test-retest reliability of CAMI by comparing the children's scores across repeated imitation trials. We demonstrate how well imitation scores from the CAMI vs HOC methods classify the children into diagnostic categories. Finally, we present CAMI's clinical significance by examining imitation performance in the ASC and TD groups and its association with core autism symptoms.

We hypothesised that while CAMI scores would highly correlate with HOC scores and show high test-retest reliability, CAMI would outperform HOC when used for distinguishing the diagnostic groups. Further, we expected that CAMI scores would yield clinically meaningful results by revealing poorer imitation in children with ASC as compared to TD children and showing strong associations with core autism symptoms.

Methods and Materials

Participants

The data reported here were collected as part of a wider-scale study examining imitation skills in autism. Our participants were 48 children (27 ASC, 21 TD) aged 8 to 12 years.

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) and parent-report of Social Responsiveness Scale (SRS-2) was 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). For all participants, ADOS-2 module-3 was used. In addition, to account for autism-associated differences in general motor abilities, we used the Movement Assessment Battery for Children, 2nd edition (mABC-2). Descriptive statistics of participant characteristics can be found in Table 1. See the Supplementary Information (SI) for full inclusion/exclusion criteria.

Ethics approval was received from [masked for blind peer-review] prior to study commencement. Written informed consent were obtained from all participants' legal guardians as well as verbal assent from all children. All recruitment took place through contacts with local schools and community events.

Procedure

Children took part in an imitation task comprised of 14 trials presented at varying movement speeds. To avoid any confound of changing movement speeds, in this study, we report only on four trials presented at 100% speed: the first two trials (Trial 1a and Trial 2a) and the last two trials (Trial 1b and Trial 2b). The last two trials were repetitions of the first two trials. These trials were of two separate movement sequences (sequence 1 = Trial 1a and Trial 1b, sequence 2 = Trial 2a and Trial 2b). The sequences comprised of 14–18 individual movement types, which were relatively unfamiliar (e.g., moving arms up and down like a puppeteer), did not have an end-goal and required moving multiple limbs simultaneously. The choice of these movement sequences was based on prior research showing particular difficulties in ASC with these types of movements [5,6,8,9,11].

The stimulus video was displayed on a large TV screen and depicted dance-like whole-body movements of a young woman without any background music/sound. The children's movements were recorded using two Kinect Xbox cameras at 30 frames per second, one located in front of the child and one at the back. Since Kinect Xbox records depth data, no sensors or special clothing was needed for this data collection. For more information about the study set-up, see SI.

The session began with a brief training phase, familiarising the participants to the kinds of movements they would perform and how much to move their bodies. All participants were instructed to perform whole-body movements and to try their best to copy the model.

Data coding

Calculation of CAMI scores—The x-y-z coordinates of 20 joints were extracted from the children's depth recordings using iPi Motion Capture Software. Children's motion data were compared to the “gold standard”, defined here as the motion data of the video model imitating herself. Imitation scores for each child were obtained following the steps outlined below. The details of the CAMI method and equations used can be found in the SI, CAMI Algorithm section.

1. Pre-processing: The child's and the gold standard's motion data are translated by locating their hips' positions at the origin. The child's limb lengths are normalised to the gold standard's skeleton, and the child's spatial orientation in the first frame is adjusted to match the gold standard.

2. Automatic joint importance estimation: Using the gold standard data, the relative contributions of each joint for each movement type are computed based on the amount of displacement observed. Joints that were displaced more in the gold standard data for a given movement type are considered to contribute more to the movement and hence affected the imitation score more than joints that stayed static.

3. Computation of the distance feature: Using DTW [17], the child's time-course is aligned to the model's time-course for the entirety of the sequence by finding a time warp that minimises the Euclidean distance between them. The DTW distances of each movement

type are calculated considering the relative importance of each joint as computed in step-2. The distances for the movement types are then averaged to make up the child's total DTW distance (s_{dist}), which is then transformed into a distance score (s_{dist}).

4. Computation of the time features: Using the DTW warping path information, time asynchrony features are computed for the entire sequence [36]: 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}) are linearly combined to make up the child's imitation score. The weights used for this linear combination are determined in a data-driven manner using 3-fold cross-validation technique. In this technique, firstly, the data set is split into 3 non-overlapping groups with equal proportions of ASC to TD children in each group. Then, two of these groups (i.e., training set) are used to learn the weights in a way that maximises the average correlation between CAMI and HOC across the trials of the training set. Using the learnt parameters, the CAMI scores of the third group (i.e., the test set) are calculated. Using cross-validation ensures that children's CAMI scores are calculated completely independently from and without reference to their HOC scores. The same procedure is repeated by assigning a new group as the test set until the CAMI scores are obtained for all three groups. The formula used for learning the weights, and the parameter values can be found in SI, Parameter Learning section.

Regarding the number of cross-validation folds, studies have shown that too few folds can lead to biased estimators, and too many folds generate high variance in the estimations [37]. Hence, we repeated the analyses using 10 folds, which replicated the same findings (see SI, Results section). Since variability was considerably larger in the 10-fold scheme (12.2%) as compared to the 3-fold scheme (3.6%), we are reporting the findings from the 3-fold scheme in the main text.

Calculation of HOC scores

To establish construct validity of the CAMI method, we analysed three trials (Trials 1a, 1b and 2a) using the more traditional HOC method. At least 40% of the videos within each trial, evenly split across diagnostic groups, were reliability-coded by two diagnosis-blind coders (all $K > .92$, $p < .001$). No HOC was done for Trial 2b videos purposely to use this trial as the replication data set.

Our HOC scheme identified the components of all movement types within a sequence (e.g., bring right arm to the right), the style of the movements (e.g., twirl arm, right/left side) and the number of repetitions. Children's total HOC score was the sum of positive items (s_{pos}) and negative items (s_{neg}) for each movement type, divided by the maximum possible score for that movement type. (s_{pos}) comprised of scores given to components successfully completed (score of +1). (s_{neg}) comprised of scores given to movements performed on the reverse side (score of -0.5) and repeated more times than demonstrated by the model (score of -1). Consequently, the children could receive a score within the range of 0–166 for sequence 1, which had 176 components and 0–202 for sequence 2, which had 216

components. These scores were then normalised to a range of 0–1, where 1 indicates perfect imitation and 0 indicates worst imitation (see SI, HOC Scheme section).

Results

To enable replications and use by future research, we provide the learnt parameters of CAMI in SI, Results. While developing the CAMI method, we proposed that it would have three main advantages to alternative methods: (1) considering temporal, in addition to spatial, differences in imitation performance, (2) assessing imitation ability with high sensitivity by yielding continuous, rather than discrete, scores, and (3) automatically detecting which joints are important for different movement types without human input. Beyond theoretical plausibility of these arguments, we conducted rigorous experiments, which empirically confirmed that these properties did indeed improve CAMI's performance (see SI, Results).

Construct validity and test-retest reliability of CAMI

To establish CAMI's construct validity, we examined its correlation with the scores obtained from HOC in three trials of movement data. The results revealed strong positive correlations between the two methods for all three trials (Figure 1a; Trial 1a: $r(43) = .82$, $p < .0001$; Trial 1b: $r(40) = .87$, $p < .0001$; Trial 2a: $r(46) = .69$, $p < .0001$). Notably, the correlation between the two methods was lowest for Trial 2a. It is worth highlighting here that the CAMI scores are calculated using a 3-fold cross-validation method, which means that children's CAMI scores were calculated independently from their HOC scores. Further supporting this point, when the same correlation tests are run between HOC scores and the distance output of DTW, which is completely unsupervised by HOC, we still observed strong correlations between the two variables (Trial 1a: $r(43) = -.78$, $p < .0001$; Trial 1b: $r(40) = -.82$, $p < .0001$; Trial 2a: $r(46) = -.70$, $p < .0001$), such that increased spatial difference between the child and the model was correlated with worse HOC scores.

We assessed CAMI's test-retest reliability by comparing performance scores between repetitions trials, comparing Trial 1a to 1b, and Trial 2a to 2b with Pearson's correlation tests. The results revealed excellent test-retest reliability (Trial 1a-1b: $r(37) = .86$, $p < .0001$; Trial 2a-2b: $r(36) = .73$, $p < .0001$).

Diagnostic classification ability of CAMI

We assessed how well imitation scores obtained from CAMI and HOC methods would classify children into diagnostic groups in two ways: (1) by training a standard machine learning algorithm (linear support vector machines, or SVM) to classify the subjects into diagnostic groups using their imitation scores as the sole features, and (2) by computing the receiver operating characteristic (ROC) curve of each trial, with larger areas under the curve (AUC) indicating better discriminative ability. Notably, the features used in SVM carry no prior information about the children's diagnosis status; the only feature used for classification was the diagnosis-blind imitation scores.

For all three trials that had both CAMI and HOC scores, the discriminative ability of CAMI was either comparable to or better than the discriminative ability of HOC. The SVM method showed that when the participants are characterised by their imitation performance across

the trials, the diagnostic groups (ASC vs TD) are more vividly separated by CAMI scores as compared to HOC scores (Figure 1b). This visible trend is supported by the higher average classification accuracy obtained by a linear SVM classifier trained in a 3-fold cross validation scheme ($\text{accuracy}_{\text{CAMI}} = 87.2\%$, $\text{accuracy}_{\text{HOC}} = 74.4\%$).

The AUC was comparable between the CAMI and HOC methods for Trials 1a and 1b, while it was considerably higher for CAMI scores in Trial 2a (Figure 1c). Considered together with the lower correlation between CAMI and HOC in Trial 2a (Figure 1a), this finding attests to CAMI's validity and superior diagnostic classification ability.

Finally, we replicated these findings in a single, one-minute imitation trial (Trial 2b), which only has CAMI scores and no HOC scores. Figure 1d shows the classification accuracy (84.6%) and AUC of CAMI based only on Trial 2b scores. Overall, these findings show that CAMI outperforms HOC in distinguishing children into diagnostic groups.

Clinical relevance of CAMI

To confirm that the imitation ability assessed by CAMI is relevant for a clinical autism sample, we examined the hypotheses that poorer imitation would be observed in ASC than in TD children, and that imitation deficits would be associated with increased autism symptoms.

We conducted a mixed-ANOVA test with diagnosis (ASC vs TD), assessment method (CAMI vs HOC), age, IQ and motor abilities (scores from the Movement Assessment Battery for Children, version 2) as the independent variables, and imitation score as the dependent variable. One TD and 2 ASC children were dropped from analysis due to violating normality assumptions (± 2 SDs from the mean); including these children did not change the findings. These statistical analyses were done using open-source R software [38].

The results revealed significant main effects of diagnosis ($F(1,23) = 13.41$, $p = .001$) and assessment method ($F(1,25) = 221.69$, $p < .0001$) as well as significant interaction effects of diagnosis*trial ($F(2,92) = 3.56$, $p = .03$), diagnosis*method ($F(1,25) = 14.80$, $p = .0007$), and trial*method ($F(1,92) = 10.55$, $p = .0001$). No other variable had a significant effect on imitation scores. Due to its relevance for our hypothesis, we further examined the diagnosis*trial interaction with Bonferroni-corrected pairwise tests. We found that within each trial and for both CAMI and HOC scores, children with ASC imitated more poorly than TD children (all $ps < .0001$; Figure 2). That no ceiling effects were observed in either group, indicates that, as reported in the literature before [5,6,8,9,11], the types of movements included in these sequences were challenging for both ASC and TD groups.

To examine the associations between imitation performance and core autism symptoms, we created composite scores by averaging the children's scores in three trials (Trials 1a, 1b, 2a). Core autism symptoms were measured by (i) parental reports of Social Responsiveness Scale (SRS-2), and (ii) the Autism Diagnostic Observation Schedule (ADOS-2) administered to children with ASC. Better imitation ability, as measured by both the CAMI and the HOC methods, was moderately and statistically significantly correlated with lower scores on the subscales of SRS-2 and the total SRS-2 scores (Table 1). Correlations between

the imitation scores and ADOS-2 scores were less strong, with the total scores reaching or approaching significance. The decreased association of CAMI scores with ADOS-2 likely stemmed from insufficient power since the ADOS-2 was administered only to the ASC group, while the SRS-2 was administered to all participants. These findings support the clinical relevance of CAMI by revealing significant links between the CAMI-assessed imitation deficits and core autism symptoms.

Discussion

In this study, we developed a method called Computerised Assessment of Motor Imitation (CAMI), which uses an automated, DTW-based algorithm, and presented its successful application to a clinical autism population to examine imitation deficits. Strong correspondence of CAMI with the standard HOC method confirmed our method's construct validity. Applying CAMI on two sets of repeated imitation trials involving two different movement sequences, we established CAMI's test-retest reliability. Further, the findings revealed that imitation ability as assessed by CAMI scores can distinguish children's clinical diagnosis (ASC vs TD) better than HOC scores. Clinical relevance of CAMI has been further confirmed with findings of CAMI-assessed poorer imitation in ASC than in TD, and a strong link between imitation deficits and core autism symptom severity.

CAMI addresses the outstanding issues with automatic assessment of human motion and imitation. The issues of heterogeneity (i.e., range of movements being virtually unlimited) and requirement for high sensitivity to detect nuances are addressed by using a continuous instead of a discrete output. The issues of high-dimensionality (i.e., involving spatial and temporal aspects of movements) and limited human supervision (i.e., lacking expert knowledge on importance of movement elements) are addressed by imposing a structure in the model that reduces the number of learnable parameters based on guidance from expert-based observations (i.e., HOC scores). Using expert knowledge to guide the features (i.e., *distance*, *t_{adv}* and *t_{delay}*) improved CAMI's interpretability, while deviance from HOC as a result of automatized learning processes improved CAMI's diagnostic discriminative ability. Finally, automatic detection of important joints enables combining high-dimensional data in a meaningful way for other movements, improving CAMI's scalability.

One advantage of CAMI is that unlike other automated methods that classify children into diagnostic groups based on broad differences in movement patterns [25,26,27], CAMI specifically measures imitation ability and provides an interpretable score indicating how well the children performed with respect to a model. Assessing imitation ability with a sensitive, automatic and objective method is important because there is robust evidence that imitation crucially impacts social bonding, learning, communication and interaction throughout development [1,2,3,30]. Since our method targets imitation ability in particular, it can be used to detect deficiencies from at least school ages onwards, and track performance during interventions designed to improve social-communicative function through imitation training.

Contrary to previous methods providing only dichotomous or categorical scores for imitation performance [15,16], CAMI produces a fine-grained, continuous score within a range of 0

(worst imitation) to 1 (best imitation). Continuous scores allow for capturing minute differences in imitation ability, which is especially important for populations with high variability such as ASC. Moreover, CAMI considers timing differences in addition to spatial differences. Given known deficits in coordinating the timing of actions in ASC [24,31], timing measures may importantly improve the assessment of autism-specific imitation impairments.

Using an SVM approach with 3-fold cross-validation, we demonstrated that CAMI scores distinguished children into diagnostic groups better than HOC scores. The CAMI method outperformed the HOC method in Trial 2a, which was the trial with lowest correlation between the two methods. Further, applied to another trial without HOC scores (Trial 2b), CAMI scores from a single, one-minute trial distinguished the children into diagnostic groups with 84.6% success. Altogether, our findings show that as compared to HOC, CAMI is more sensitive in detecting autism-associated differences in imitation performance.

Given the heterogeneity of the behavioural phenotypes in the autism spectrum, the sample size of 27 ASC and 21 TD children can be considered relatively small. While such inherent limitations of a small sample size should be born in mind while interpreting the findings, it is important to clarify that the CAMI machine learning approach does not suffer from a sample size problem in either the calculation of the CAMI scores or SVM classification. Firstly, since estimating the minimum sample size needed by canonical correlation analysis (a general case of the method we used to learn the CAMI scores) is non-trivial, studies suggest the “one-in-ten rule”, which states that 10 samples per variable is enough to estimate the parameters [39, 40]. In our case, when we maximise the correlation, we are working with 3 variables, thus 30 samples should suffice. In the 3-fold cross-validation scheme, we are using approximately 90 samples (2/3 of subjects, 3 trials) to estimate 3 variables. Secondly, for binary max-margin linear classifiers such as the SVMs we used, Raudys (1997) [41] provides a formula to estimate the mean expected classification error in terms of the number of parameters, distance between classes, and sample size. Applying equation 12 from Raudys (1997) to our problem, in which $p = 3$ (number of parameters), $\delta = 3.83$ (the distance between normalised class centres), and $N = 13$ (the approximate number of samples per class for training in the 3-fold cross-validation scheme), we obtain a mean expected classification error of 5.9%. Given that the minimum possible mean expected classification error for this problem (i.e., if we had infinite samples) would be 2.8%, our error rate can be considered sufficiently good. Notably, the parameter-to-sample-size ratio used here is higher than previous applications of machine learning to distinguish motor patterns in autism. For example, Li and colleagues (2017) trained a model with 40 parameters using data from 30 subjects [26], and Crippa and colleagues (2015) trained a model with 7 parameters using data from 30 subjects [27].

Future research is needed to improve the scalability of CAMI. At present, this method uses data obtained from Kinect Xbox depth cameras, which, due to imperfections of the motion tracking technology, require some manual processing that can be time consuming. Future research should explore the use of this method on 2D data obtained from high-resolution cameras. Moreover, in order for CAMI to be used widely as a clinical assessment tool, we need to establish norm-standardisation with larger data sets, including younger age groups

and varied demographics. Relatedly, given the relatively small sample size of this study for the highly heterogeneous autism population, it is crucial that the current findings be replicated in future research; administering ADOS-2 to the entire population would be informative in future replication attempts.

The CAMI method presented here is a major step forward in examining motor imitation automatically without requiring extensive human input or coder training. This method provides an objective, continuous, highly scalable and directly interpretable score. As such, CAMI can be used in clinics and home settings to assess imitation ability, to help inform diagnostic decision-making based on the children's imitation performance (e.g., autism vs non-ASC) and to advance biomarkers-based interventions for improving social-communicative functioning through imitation-based strategies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Recruitment for this study was supported by National Institutes of Health (NIH) R01 grant MH106564-02, for which [the senior author] is a co-investigator, while algorithm development was supported by NIH R01 grant HD87133-01, for which [the second senior author] is a co-investigator.

References

- [1]. Prinz W “Experimental approaches to imitation” in *The imitative mind: Development, evolution and brain bases*, Meltzoff AN & Prinz, Eds. (Cambridge: Cambridge University Press, 2002), pp. 143–162
- [2]. Heyes C, “What can imitation do for cooperation?” In *Life and mind: Philosophical issues in biology and psychology. Cooperation and its evolution*, Sterelny K, Joyce R, Calcott B, & Fraser B, Eds., (Cambridge, MA, US: The MIT Press, 2013), pp. 313–331.
- [3]. Nadel J, “Does imitation matter to children with autism?” In *Imitation and the social mind: Autism and typical development*, Rogers SJ & Willians JHG, Eds., (New York, NY, US: The Guilford Press, 2006), pp. 118–137.
- [4]. Edwards LA. A meta-analysis of imitation abilities in individuals with autism spectrum disorders. *Aut. Res* 7(3), 363–380 (2014).
- [5]. Gowen E, Imitation in autism: Why action kinematics matter. *Front. in Int. Neurosci.* 6, 1–4 (2012).
- [6]. MacNeil LK and Mostofsky SH, Specificity of dyspraxia in children with autism. *Neuropsychology*, 26(2), 165–171 (2012). [PubMed: 22288405]
- [7]. Nebel MB, Eloyan A, Nettles CA, Sweeney KL, Ament K, Ward RE, Choe AS, Barber AD, Pekar JJ and Mostofsky SH, Intrinsic visual-motor synchrony correlates with social deficits in autism, *Bio. Psych.* 79(8), 633–641 (2016).
- [8]. McAuliffe D, Pillai AS, Tiedemann A, Mostofsky SH & Ewen JB Dyspraxia in ASD: Impaired coordination of movement elements. *Autism Res* 10, 648–652 (2017). [PubMed: 27653620]
- [9]. Hobson PR and Hobson JA, Dissociable aspects of imitation: A study in autism. *J. of Exp. Child Psych.* 101, 170–185 (2008).
- [10]. Vivanti G, Nadig A, Ozonoff S, Rogers SJ, What do children with autism attend to during imitation tasks? *J. of Exp. Child Psych.* 101, 186–205 (2008).
- [11]. Marsh L, Pearson A, Ropar D and Hamilton A, Children with autism do not overimitate. *Current Biology*, 23(7), R266–268 (2013). [PubMed: 23578869]

- [12]. Carpenter M, Tomasello M and Savage-Rumbaugh S, Joint attention and imitative learning in children, chimpanzees, and enculturated chimpanzees, *Soc. Dev.* 4(3), 217237 (1995).
- [13]. Zheng Z, Das S, Young EM, Swanson A, Warren Z and Sarkar N, Autonomous robot-mediated imitation learning for children with autism, *IEEE Int. Conf. on Robotics and Automation*, (2014). doi: 10.1109/ICRA.2014.6907247
- [14]. Zheng Z, Young EM, Swanson AR, Weitlauf AS, Warren ZE and Sarkar N, Robot-mediated imitation skill training for children with autism, *IEEE Trans. on Neural Syst. and Rehab. Eng.* 24(6), 682–691 (2016).
- [15]. Michelet S, Karp K, Delaherche E, Achard C and Chetouani M, “Automatic imitation assessment in interaction”, in *Human Behavior Understanding. HBU 2012. Lecture Notes in Computer Science*, Salah AA, Ruiz-del-Solar J, Meriçli Ç, Oudeyer PY, Eds., (Springer, Berlin, Heidelberg, 2012), vol 7559, doi: 10.1007/978-3-642-34014-7_14.
- [16]. Su C-J, Chiang C-Y and Huang J-Y, Kinect-enabled home-based rehabilitation system using Dynamic Time Warping and fuzzy logic, *Applied Soft Computing*, 22, 652–666 (2014).
- [17]. Sakoe H, and Chiba S, Dynamic programming algorithm optimization for spoken word, *IEEE Transac. on Acoustics, Speech, and Signal Proc.* 26(1), 43–49 (1978).
- [18]. Reyes M, Dominguez G and Escalera S. Feature weighting in dynamic time warping for gesture recognition in depth data, *IEEE Int. Conf. on Computer Vision Workshops*, 1182–1188 (2011). doi: 10.1109/ICCVW.2011.6130384
- [19]. Celebi S, Aydin AS, Temiz TT and Arici T, Gesture recognition using skeleton data with weighted dynamic time warping, *Int. Conf. on Computer Vision Theory and App.*, (2013), doi: 10.5220/0004217606200625.
- [20]. Ofli F, Chaudhry R, Kurillo G, Vidal R and Bajcsy R. Sequence of the most informative joints (SMIJ): A new representation for human skeletal action recognition. *J. of Visual Comm. and Image Rep.* 25(1), 24–38 (2014).
- [21]. Arici T, Celebi S, Aydin AS and Temiz TT, Robust gesture recognition using feature pre-processing and weighted dynamic time warping, *Multimedia Tools and App.* 72(3), 3045–3062 (2014).
- [22]. Delaherche E, Boucenna S, Karp K, Michelet S, Achard C and Chetouani M, “Social coordination assessment: Distinguishing between shape and timing” in *Multimodal Pattern Recognition of Social Signals in Human-Computer-Interaction*, Schwenker F, Scherer S, Morency LP, Eds., (Springer, Berlin, Heidelberg, 2012), doi: 10.1007/978-3-642-3-7081-6_2.
- [23]. de Marchena A and Eigsti I-M, Conversational gestures in autism spectrum disorders: Asynchrony but not decreased frequency, *Aut. Res.* 3(6), 311–322 (2010).
- [24]. Marsh KL, Isenhower RW, Richardson MJ, Helt M, Verbalis AD, Schmidt RC and Fein D, Autism and social disconnection in interpersonal rocking, *Front. in Int. Neurosci.* 7(108) (2013).
- [25]. Guha T, Yang Z, Grossman RB and Narayanan SS. A computational study of expressive facial dynamics in children with autism, *IEEE Transac. on Affective Computing*, 9(1), 14–20 (2015).
- [26]. Li B, Sharma A, Meng J, Purushwalkam S and Gowen E, Applying machine learning to identify autistic adults using imitation: An exploratory study, *PloS ONE*, 12(8), (2017).
- [27]. Crippa A, Salvatore C, Perego P, Forti S, Nobile M, Molteni M and Castiglioni I, Use of machine learning to identify children with autism and their motor abnormalities, *J. of Aut. and Dev. Dis.* 45(7), 2146–2156 (2015).
- [28]. Fawcett C and Liszkowski U, Mimicry and play initiation in 18-month-old infants, *Infant Beh. and Dev.* 35(4), 689–696 (2012).
- [29]. Carpenter M, Uebel J and Tomasello M, Being mimicked increases prosocial behavior in 18-month-old infants, *Child Dev.* 84(5), 1511–1518 (2013). [PubMed: 23488734]
- [30]. Stel M and Vonk R, Mimicry in social interaction: Benefits for mimickers, mimicked, and their interaction, *British J. of Psych.* 101, 311–323 (2010).
- [31]. Oberman LM, Winkielman P and Ramachandran VS, Slow echo: Facial EMG evidence for the delay of spontaneous, but not voluntary, emotional mimicry in children with autism spectrum disorders, *Dev. Sci* 12(4), 510–520 (2009). [PubMed: 19635079]

- [32]. McIntosh DN, Reichmann-Decker A, Winkielman P and Wilbarger JL, When the social mirror breaks: Deficits in automatic, but not voluntary, mimicry of emotional facial expressions in autism, *Dev. Sci* 9(3), 295–302 (2006). [PubMed: 16669800]
- [33]. Field T, Field T, Sanders C and Nadel J, Children with autism display more social behaviors after repeated imitation sessions, *Autism*, 5(3), 317–323 (2001). [PubMed: 11708590]
- [34]. Contaldo A, Colombi C, Narzisi A and Muratori F, The social effect of “being imitated” in children with autism spectrum disorder, *Front. in Psych*, 7, 1–16 (2016).
- [35]. Heimann M, Laberg KE and Nordøen B, Imitative interaction increases social interest and elicited imitation in non-verbal children with autism, *Infant and Child Dev*, 15, 297–309 (2006).
- [36]. Folgado D, Barandas M, Matias R, Martins R, Carvalho M and Gamboa H, Time alignment measurement for time series, *Pattern Recognition*, 81, 268–279 (2018).
- [37]. Rodríguez JD, Pérez A, Lozano JA, Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* 32, 569–575 (2010). [PubMed: 20075479]
- [38]. R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2019) URL <https://www.R-project.org/>.
- [39]. Weinberg SL, Darlington RB, Canonical analysis when number of variables is large relative to sample size. *J. Educ. Stat.* 1, 313–332 (1976).
- [40]. Tabachnick BG, Fidell LS, *Using Multivariate Statistics* (Pearson Education, Seventh., 2019).
- [41]. Raudys S, On dimensionality, sample size, and classification error of nonparametric linear classification algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* 19, 667–671 (1997).

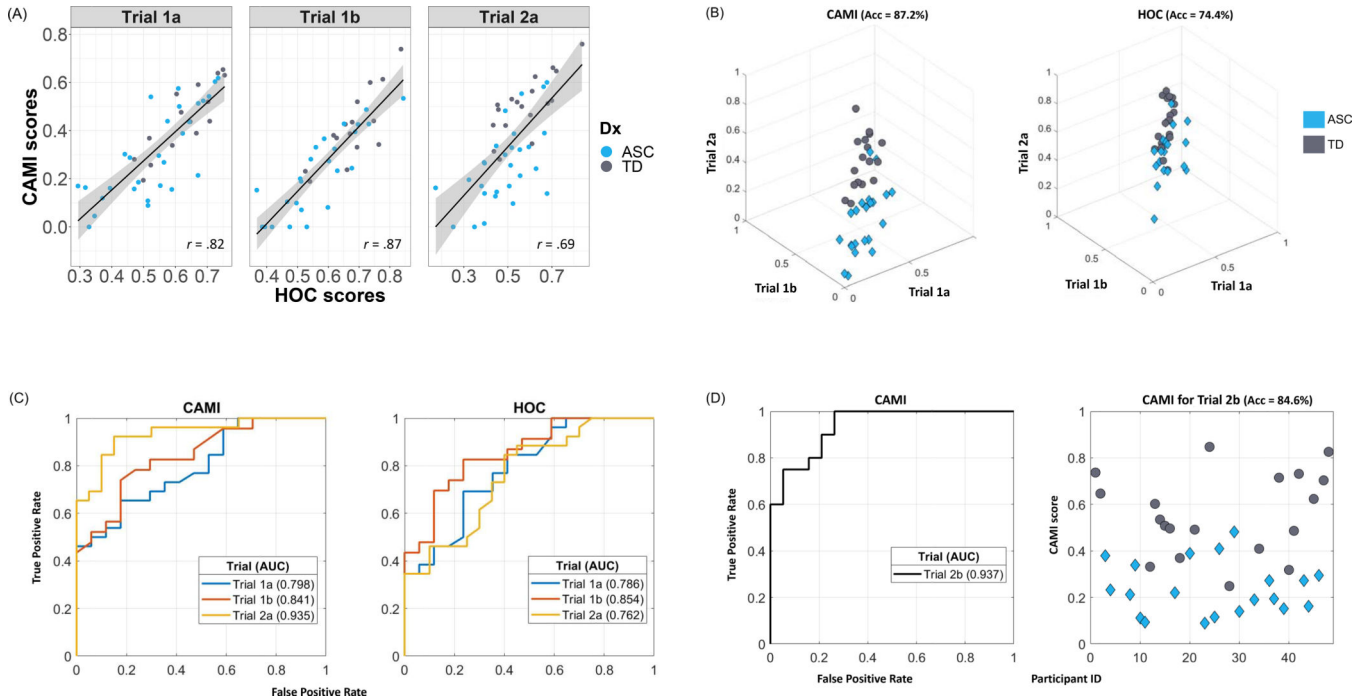


Figure 1. Comparisons between the CAMI and HOC methods using motion data of four imitation trials from a sample of 48 children (27 ASC, 21 TD).
 (a) Correlations between the CAMI scores and HOC scores in three trials, showing strong correspondence between the two methods. An $r = 1$ indicates perfect positive association, $r = 0$ indicates no association and $r = -1$ indicates perfect negative association ($p < .0001$).
 (b) 3D plots of the CAMI scores and HOC scores in which Trial 1a, Trial 1b, and Trial 2a scores correspond to the respective axes. Each marker represents one subject, and the reported accuracy (Acc) corresponds to average classification accuracy in 3-fold cross-validation of a linear SVM classifier (best possible Acc is 100%, meaning all participants categorised to diagnostic groups accurately).
 (c) Receiving Operating Characteristic (ROC) curves: true positive rate versus false positive rate as classification threshold is varied. The Area Under the Curve (AUC) indicates the diagnostic ability of the method (left panel for CAMI, right panel for HOC) in each of the three trials (best possible AUC is 1, meaning zero false positives and 100% true positives).
 (d) ROC curve (left panel) and CAMI scores (right panel) of Trial 2b only. Since this trial does not have any HOC scores, its CAMI scores are computed based on parameters learnt from the other three trials, complying with the splits used for 3-fold cross-validation. The AUC of the ROC curve (0.937) and SVM accuracy value (84.6%) demonstrate the diagnostic classification ability of CAMI scores with this single trial.

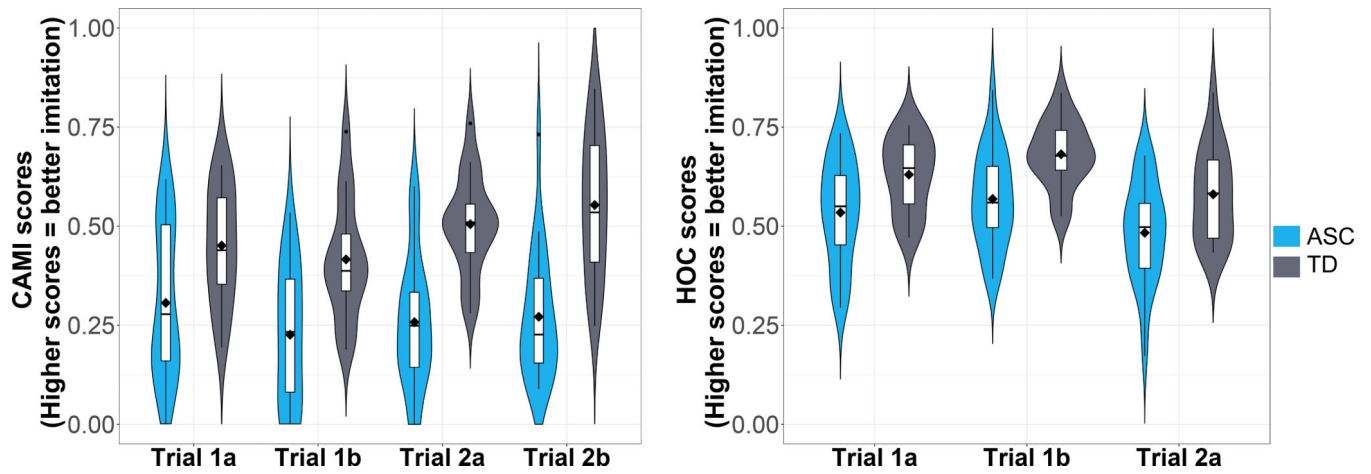


Figure 2.

Imitation performance per diagnostic group (blue = ASC, grey = TD) per trial according to CAMI scores (left) and HOC scores (right) with box plots embedded within violin plots. In the box plots, horizontal lines indicate medians, boxes indicate data within the 25th to 75th percentiles, and whiskers indicate data within the 5th to 95th percentiles.

Participant characteristics. Here, $t(N)$ and $\chi^2(n,N)$ denote, respectively, the student-t and chi-square statistics, where N is the sample size and n is the number of degrees of freedom.

Table 1.

| | ASC | | TD | | Test statistics (ASC vs TD) |
|----------------------------------|--------------|------------|---------------|------------|------------------------------------|
| | Mean (SD) | Range | Mean (SD) | Range | |
| Chronological age (years) | 10.34(1.42) | 8.03–12.83 | 10.41(1.26) | 8.55–12.73 | $t(46) = -1.9$ $p > .05$ |
| SRS-2 total score | 75.73(7.40) | 60–87 | 44.81(4.04) | 39–54 | $t(40) = 15.33$ $p < .0001$ |
| ADOS-2 total score | 15.54(4.34) | 8–27 | — | — | — |
| WISC-V Full scale IQ | 98.41(15.62) | 70–130 | 109.86(12.06) | 94–143 | $t(46) = -2.78$ $p = .008$ |
| Gender (Boys/Girls) | 24/3 | — | 18/3 | — | $\chi^2(1,48) = 0.11$ $p > .05$ |

Table 2.

Correlation between imitation ability and core autism symptoms, denoted as $r(N)$, where N is the sample size, $r = 1$ indicates perfect positive association, $r = 0$ indicates no association, and $r = -1$ indicates perfect negative association. SCI: Social Communication and Interaction, RRB: Restricted Interests and Repetitive Behaviour, SA: Social Affect. Higher scores in SRS-2 and ADOS-2 indicate greater autism severity.

| Imitation ability | Social Responsiveness Scale - 2 (parent report) | | | Autism Diagnostic Observation Scale - 2 (in the ASC group only) | | |
|--------------------|--|----------------------|------------------|--|------------------|---------------------------|
| | SCI sub-scale | RRB sub-scale | Total Score | SA sub-scale | RRB sub-scale | Total Score |
| CAMI method | $r(40) = -.48^*$ | $r(40) = -.52^{***}$ | $r(40) = -.50^*$ | $r(24) = -.17$ | $r(24) = -.23^*$ | $r(24) = -.34^{\ddagger}$ |
| HOC method | $r(40) = -.42^*$ | $r(40) = -.45^*$ | $r(40) = -.43^*$ | $r(24) = -.39^{\ddagger}$ | $r(24) = -.21$ | $r(24) = -.42^*$ |

‡ $p < .10$

$*$ $p < .05$

*** $p < .001$