

# Can low-cost motion-tracking systems substitute a Polhemus system when researching social motor coordination in children?

Veronica Romero<sup>1</sup> · Joseph Amaral<sup>1,2</sup> · Paula Fitzpatrick<sup>3</sup> · R. C. Schmidt<sup>4</sup> · Amie W. Duncan<sup>2</sup> · Michael J. Richardson<sup>5</sup>

Published online: 29 April 2016  
© Psychonomic Society, Inc. 2016

**Abstract** Functionally stable and robust interpersonal motor coordination has been found to play an integral role in the effectiveness of social interactions. However, the motion-tracking equipment required to record and objectively measure the dynamic limb and body movements during social interaction has been very costly, cumbersome, and impractical within a non-clinical or non-laboratory setting. Here we examined whether three low-cost motion-tracking options (Microsoft Kinect skeletal tracking of either one limb or whole body and a video-based pixel change method) can be employed to investigate social motor coordination. Of particular interest was the degree to which these low-cost methods of motion tracking could be used to capture and index the coordination dynamics that occurred between a child and an experimenter for three simple social motor coordination tasks

in comparison to a more expensive, laboratory-grade motion-tracking system (i.e., a Polhemus Latus system). Overall, the results demonstrated that these low-cost systems cannot substitute the Polhemus system in some tasks. However, the lower-cost Microsoft Kinect skeletal tracking and video pixel change methods were successfully able to index differences in social motor coordination in tasks that involved larger-scale, naturalistic whole body movements, which can be cumbersome and expensive to record with a Polhemus. However, we found the Kinect to be particularly vulnerable to occlusion and the pixel change method to movements that cross the video frame midline. Therefore, particular care needs to be taken in choosing the motion-tracking system that is best suited for the particular research.

**Keywords** Motion-tracking · Social motor coordination · Coordination dynamics · Microsoft Kinect · Polhemus Latus system

✉ Veronica Romero  
romerovc@mail.uc.edu

✉ Michael J. Richardson  
richamo@ucmail.uc.edu

<sup>1</sup> Center for Cognition, Action and Perception, Department of Psychology, University of Cincinnati, Edwards C1, 4146, Cincinnati, OH 45221, USA

<sup>2</sup> Department of Developmental and Behavioral Pediatrics, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati, OH, USA

<sup>3</sup> Department of Psychology, Assumption College, Worcester, MA, USA

<sup>4</sup> Department of Psychology, College of the Holy Cross, Worcester, MA, USA

<sup>5</sup> Center for Cognition, Action and Perception, Department of Psychology, University of Cincinnati, Edwards C1, 4150B, Cincinnati, OH 45221, USA

Functionally stable and robust interpersonal motor coordination, including both behavioral synchrony and imitation, has been found to play an integral role in the effectiveness of social interactions (e.g., Bernieri & Rosenthal, 1991; Chartrand & Bargh, 1999; Marsh et al., 2013; Knoblich & Sebanz, 2006; Sebanz & Knoblich, 2009). However, the motion-tracking equipment required to record and objectively measure the dynamic limb and body movements of children (or even adults) during social interaction has been very costly, as well as cumbersome and impractical within a non-clinical or non-laboratory setting. Thankfully, over the last 5 years (or so) an increasing number of low-cost motion-tracking systems (e.g., Microsoft Kinect, Microsoft LTD), or alternative video-based methods (e.g., Paxton & Dale, 2013) of motion capture have become available to researchers and clinicians interested

in investigating the behavioral dynamics of human motor control and social motor coordination. A result of the rapid advancements in the computational power of modern computers that has occurred over the last decade and the increase in consumer demand for more interactive-based video-game systems, many of these systems are able to track human movements wirelessly (i.e., remotely), with few-to-no sensors attached to the participant, and with nominal computer equipment (i.e., can be performed using almost any modern laptop computer). In addition to costing only a fraction of the price of their high-end laboratory standard counterparts, these systems are easy to replace, highly portable, and can be used almost anywhere (i.e., both clinical/laboratory and non-clinical/non-laboratory settings). Furthermore, they typically come with companion open-source software or software development kits that enable researchers to develop applications, testing protocols, and data analysis systems that meet the specific needs of the researcher or research population in question.

The degree to which these systems are able to replace more expensive laboratory-grade motion-tracking systems for research on social motor coordination in children, adults, and special populations is therefore an important question that needs to be addressed. It is likely that the use of such systems will be task and behavior dependent, and that the difference in their spatial and temporal accuracy compared to laboratory-grade systems will significantly influence the types of methodologies or research protocols that can be employed. Of particular interest is the degree to which research findings obtained with these low-cost systems can be considered reliable.

To explicate the viability of these low-cost systems for investigating social motor coordination, we conducted a study investigating social motor control in typically developing children (TD) and children with autism spectrum disorder (ASD), using four methods of motion capture: (a) a high-end (yet moderately affordable) laboratory-grade Polhemus Latus magnetic motion-tracking system, (b) the Microsoft Kinect motion-tracking sensor, which is a low-cost optical tracking system, from which we extracted the limb movements, (c) Microsoft Kinect from which we can extract whole body movement, and (d) a video recording-based pixel-change method of whole-body motion extraction. Additionally, we investigated three different types of social movement tasks that vary in the kind of motor coordination (in-phase, anti-phase) they exhibited as well as how much a single limb or the whole body was used in the task. Below, we provide a brief description of these different methods and a detailed comparison of how these low-cost methods of motion capture fared with respect to determining the stability and patterning of the social coordination that occurred across a range of interpersonal motor tasks. Of particular interest was how well the low-cost Microsoft Kinect and video pixel-change methods performed in comparison to the more expensive, laboratory-grade, Polhemus Latus system. Specifically, we tested the degree to which these three methods

could be employed to (a) differentiate the type of coordination that occurred in these tasks, and (b) capture the coordination differences evident in participants with and without ASD.

## Motion capture systems and methods employed

### Polhemus Liberty Latus wireless

This motion-tracking system is a wireless motion-tracking system developed by Polhemus LTD (Vermont, USA; <http://www.polhemus.com>; Liberty Latus Brochure, 2012) that uses an electromagnetic field to map the position (Euclidian  $x$ ,  $y$ , and  $z$  coordinates) and rotation (pitch, yaw, roll) of 1–12 small 79.4-g sensors/markers. The system tracks these six-degrees-of-freedom sensors within an electromagnetic capture volume that is defined by a map of 1–16 receptors. Each receptor has an optimal diametric capture volume of 6 f. and multiple sensors can be aligned by the user (experimenter/clinician) to meet the spatial demands of the behavior(s) performed or recording volume required. That is, the position and rotation of a sensor is tracked so long as it is in range of at least one receptor within the map of receptors that defines the capture volume setup by the user – the more receptors, the larger the possible capture volume (multiple systems can also be daisy-changed for even larger volumes). The reliability and resolution of this equipment is excellent, with a sampling rate of 188 Hz or 94 Hz (i.e., samples/s) and a positional and rotational resolution of approximately 0.25 cm and 0.5° (if a marker/sensor is no more than 4 f. from a receptor). Although the system does require a significant amount of time to set up the capture volume (i.e., the receptor volume), once the capture volume has been defined the system is easy to use and can be used with multiple participants with little to no calibration. Unlike optical tracking systems (such as an OptoTrak or Vicon systems), the Polhemus Latus is not susceptible to occlusion and can therefore be used for almost any motor task and in any environment as long as it does not include metal or electronic devices that might interfere with the electromagnetic field defined by the receptors. The system costs approximately US\$12,500.00 for a one-marker/one-receptor system and approximately US\$60,000.00 for a 12-marker/16-receptor system.

### Microsoft Kinect

The Kinect sensor (version 1)<sup>1</sup> combines a specialized video camera and an infrared depth-sensing emitter to optically track the Euclidian  $x$ ,  $y$ , and  $z$  location (in coordinates relative to sensor placement) of up to 21 skeletal/body joints (i.e., head,

<sup>1</sup> Since completing this study, Microsoft released a new version of the Kinect Sensor (i.e., version 2). Although this new version has improved voice and person recognition features, the temporal and spatial resolution of skeletal (motion capture) tracking has remained the same. Thus the current results should generalize to the Kinect Sensor Version 2.

left/right shoulders, elbows, wrists, the spine, left/right hips, knees, feet,... etc.; Kinect for Windows Sensor Components and Specifications, 2014). The device was originally developed by Microsoft for their Xbox gaming console, but can also be purchased for use on any PC or laptop computer running a Windows XP operating system and above. The research version costs approximately US\$225.00 and is capable of capturing skeletal/joint data and color BMP/video images at a maximum rate of 30 Hz (i.e., 30 frames/s), with a resolution of 1,280 × 960 pixels. A free C/C++ and C# SDK is available directly from Microsoft and can be used to develop non-commercial applications and recording software. The Kinect system is completely wireless and does not require any sensors to be placed on the body of the individual being tracked (which makes it especially useful when collecting data from children with ASD). However, since the skeletal data is based on a combined infrared/video process of depth and a machine-learning algorithm trained extensively with the use of synthetic depth images for its inference of motion tracking (Shotton et al., 2011), it requires a constant line of sight of the limbs/bodies being tracked and is especially susceptible to occlusion. It also has a high noise-to-signal ratio (relative to the Polhemus Latus system, for example), such that it is typically unable to reliably capture small or subtle changes in limb or body position, especially when participants are wearing loose clothing or the system is used in a high UV lighted environment.

### Video pixel change motion extraction

This method of motion analysis involves calculating the amount of pixels that change between adjacent video frames, which can be taken to index the amount of activity of a participant if they are the only source of movement in that part of the frame (Kupper et al., 2010; Paxton & Dale, 2013; Schmidt et al., 2012). This calculation process can be automated using simple video analysis routines written in Matlab (Mathworks, Inc., Natick, MA) or similar data analysis and scripting software, and can even be employed to extract the global movement of two (or more) individuals so long as their movements or activity are within the same recorded frame. That is, video frames can be cropped to include the movements of only one person (i.e., the left half or right half of the screen) and also the absolute difference of pixel change between the adjacent frames of the video when calculated to form an image-change time series for each participant in the interaction (see below for more details).

## Experimental method

### Participants

Thirty-eight children (seven female) between 6 and 10 years of age were recruited to participate in the study. Nineteen

typically developing children (M age = 8.15, SD age = 1.34) and 19 children that had been diagnosed with ASD (M age = 7.84, SD age = 1.46). The diagnosis of these children was corroborated within 3 months of the experimental session as part of the study with the administration of the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2), which also provides a measure of severity. All of the ASD participants were considered to be high functioning. None of the children placed in the typically developing group had ever received a psychological diagnosis.

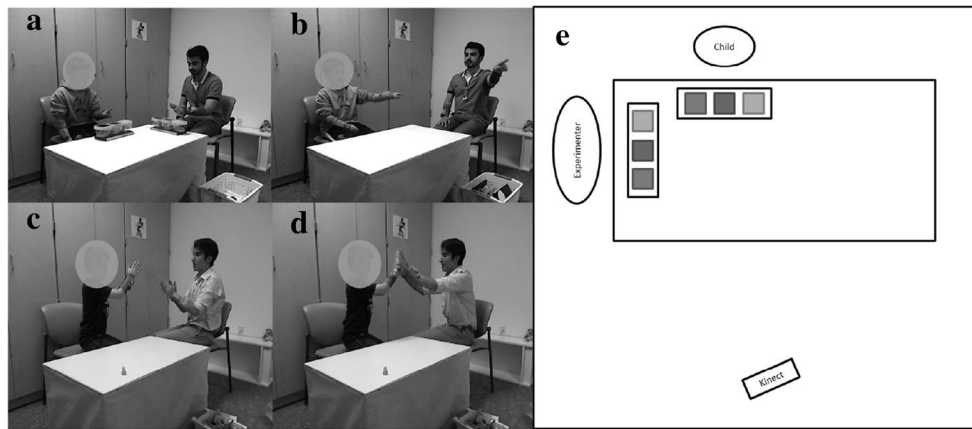
### Equipment setup

The study was conducted in a 10 × 12 f. laboratory room at Cincinnati Children's Hospital Medical Center (University of Cincinnati, Cincinnati, OH, USA). Children came into the laboratory room and were asked to sit at a 2-ft wide × 4-ft long × 2-ft high table next to the seated experimenter (see Fig. 1). Four Polhemus Latus receptors were attached to the underside of the table top, one in each corner, to create a 10 × 12 × 8 f. capture volume around the table. As soon as the child was seated, the four Polhemus Liberty Latus wireless markers/sensors were placed in wristbands and slipped over the child's and experimenter's wrists (one marker on each wrist of the child and experimenter). The motion of the Polhemus sensors was recorded at 94 Hz on a Dell PC computer using a custom software application written by the authors using the Polhemus Latus C/C++ SDK Library.

The Microsoft Kinect sensor was placed at a height of 1.5 m, 3 m away from corner of the table top closest to the participant and experimenter at approximately a 45° angle (see Fig. 1e). A custom software application ([www.xkiwilabs.com](http://www.xkiwilabs.com)) using the free Windows Kinect SDK version 1.5 (Microsoft LTD) was used to record the head, spine, and upper body skeletal data (11 skeletal points in total; no hip, leg, or foot data were recorded) of the seated child and experimenter at a sample rate of approximately 30 Hz, as well as the video images used for the pixel change analysis (also at an approximate image rate of 30 Hz).

### Coordination tasks

The data presented here were part of a larger project, in which participants performed a variety of motor, social, and cognitive tasks. Here, we selected three social motor coordination tasks that were performed by all of the children. These tasks were selected because they represent a range of movement dimensions that varied in terms of the number of effectors used (a single limb or the whole body) and the size of the movements (small or large scale). The first coordination task involved a sequence of *tapping movements*, which involved children using a finger from one hand to tap/hit three drum-like cylinders from left to right in synchrony with the



**Fig. 1** Room set-up for (a) the object tapping, (b) the pointing, and (c, d) the hand clapping game. (e) Schematic representation of the experimental room

experimenter (see Fig. 1a). Children repeated this left-to-right drumming sequence six times with the experimenter in a continuous manner. This first task was done with a single limb and involved small-scale movements. The second task involved a sequence of *pointing movements*, in which children were required to point at approximately shoulder height to the right, center, and left of their body midline in synchrony with the experimenter (see Fig. 1b). Again, children repeated this pointing sequence six times with the experimenter in a continuous manner. The second task involved the use of only one limb, but in this case the movements were larger scale. The third task was an interpersonal hand clapping game (a version of the pat-a-cake game), in which children completed a simple repetitive sequence of clapping their left and right hands together and then with the experimenter (see Fig. 1c and d). The hand clapping game was completed twice, with each sequence consisting of six consecutive intrapersonal and interpersonal clapping movements; however, the data presented here is only from the second trial of the hand clapping game.<sup>2</sup> This task in turn involved whole-body, larger-scale movements.

### Motion data reduction

All the data extraction and analysis methods presented below were completed using custom MATLAB (Mathworks, Inc., Natick, MA) applications and functions developed by the authors. These MATLAB applications and functions, as well as example time series can be downloaded from [www.xkiwilabs.com](http://www.xkiwilabs.com).

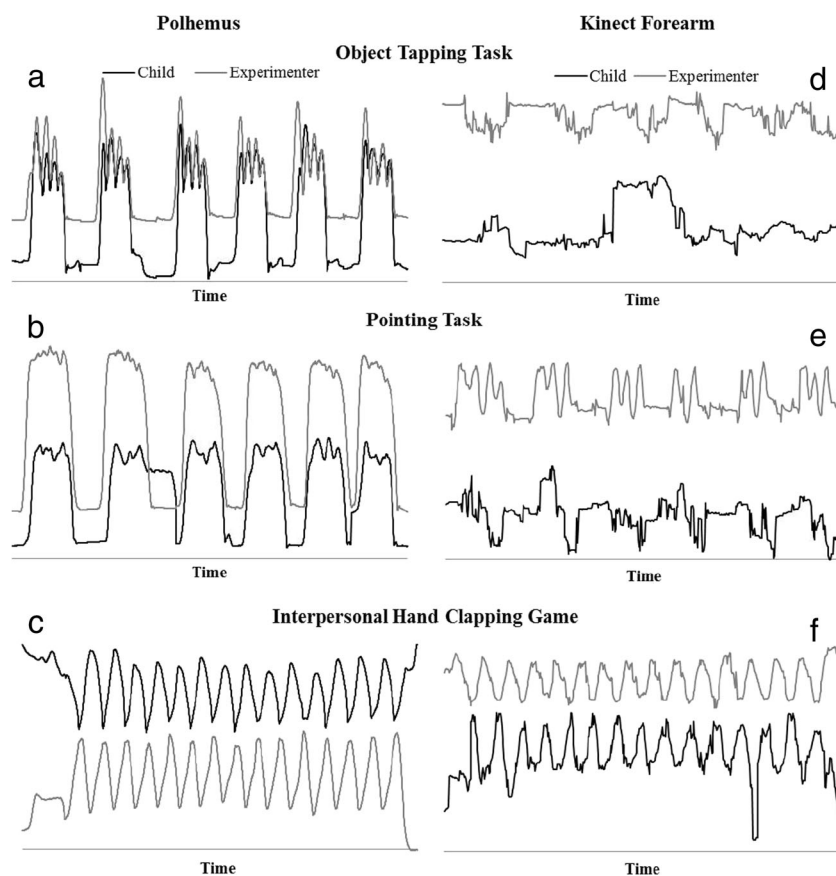
<sup>2</sup> Due to the experiment design, an experimenter walked in front of the Kinect sensor before the start of the hand clapping game which led to a lack of skeletal tracking through the Kinect for the first trial for some participants. Consequently, we decided to use only the second trial. In fact, skeletal tracking was completely missing for the second trial of two participants. Hence, their data were excluded from the result section below.

**Polhmeus Latus** The  $x$ -plane (left-right),  $y$ -plane (forward-back), and  $z$ -plane (up-down) positional coordinates of the sensors placed on the wrists of the experimenter and child were recorded for each task. To best determine the stability and patterning of the behavioral coordination that occurred between the child and experimenter, we first isolated the primary plane of motion for each task. Since the primary plane of motion for the drumming and pointing tasks was in the left-right plane, the  $x$ -plane movement time series was used to assess the behavioral coordination that occurred for these two tasks (for a sample time series see Fig. 2a and b). For the interpersonal hand clapping game, the largest amplitude of movement was in the up-down,  $z$ -plane, with the intrapersonal clapping events occurring at a lower height than the interpersonal clap events (for a sample time series see Fig. 2c). Accordingly, this plane of motion was employed to assess the behavioral coordination that occurred for this task.<sup>3</sup>

For the tapping and pointing tasks, we then performed a coordination analysis (see below for details) using the primary plane of motion time series of the experimenter's right forearm (the experimenter always used his right hand/arm for all the tasks) and the primary plane of motion time series of the forearm used by the child for analysis. Note that for the tapping and pointing tasks the child was free to use either left or right arm/hand. Although both arms/hands were employed by the experimenter and child for the hand clapping game, we only analyzed the right forearm movements of the experimenter and child because the coordination that occurred between the left forearm movements was completely redundant with the right.

**Microsoft Kinect** The data recorded from the Kinect was extracted for analysis using two different methods. The first method, which will be referred to as the Kinect

<sup>3</sup> An analysis of secondary planes of motion for the different methods of motion capture being discussed produced results that were consistent with those reported here



**Fig. 2** Sample movement time series from a child collected by the Polhemus Latus in (a) the object tapping task, (b) the pointing task, and (c) the interpersonal hand clapping game. Also, sample movement time

series from the same child collected by the Kinect forearm in (d) the object tapping task, (e) the pointing task, and (f) the interpersonal hand clapping game

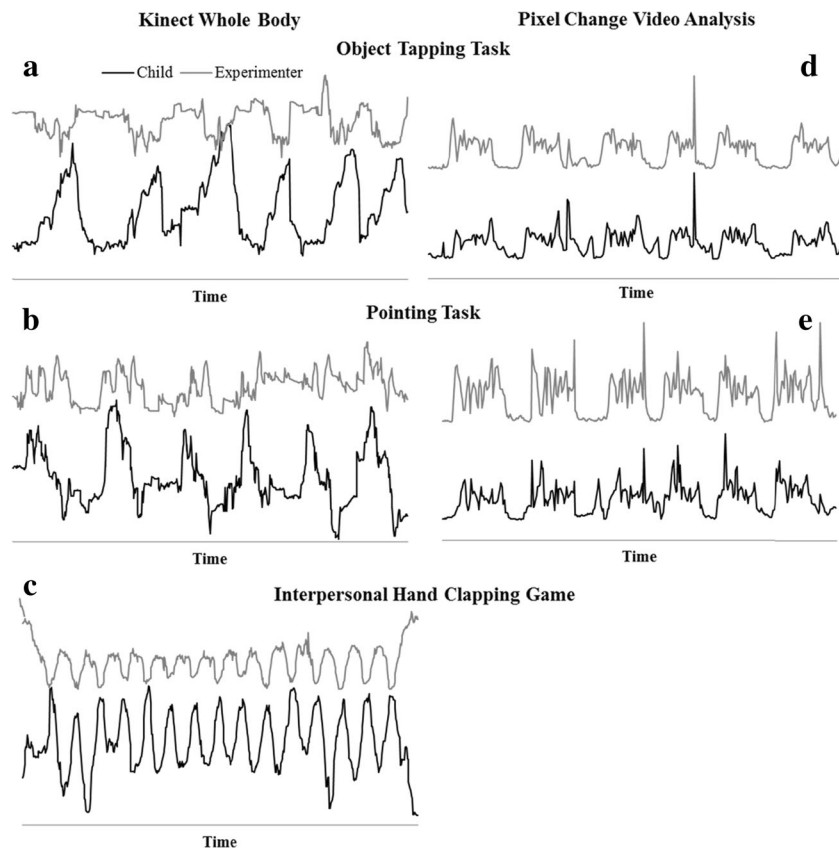
forearm method, was comparable to the method used for the Polhemus Latus system described in the preceding section. That is, the child's and experimenter's forearm movements in the  $x$ -,  $y$ -, and  $z$ -planes were extracted for the tapping, pointing, and hand clapping game, and an additive time series was created (for examples of these time series see Fig. 2d, e, and f). The same forearm-side combinations identified for the Polhemus Latus analysis were also employed.

The second method, which will be referred to as the Kinect whole body method, involved creating a unified one-dimensional movement time series for both the child and experimenter from the  $x$ -,  $y$ -, and  $z$ -plane motion of all of the upper-body joints recorded by the Kinect sensor (i.e., the spine, head, and the left and right shoulder, elbow, hand, and wrist). This was achieved by simply creating a vector based on the sum of the values of each movement/joint dimension at each time-step (for examples of these time series see Fig. 3a, b, and c). This method of normalization was chosen in order to produce a “collective” whole body motion time series for the child and experimenter that would be similar to the collective motion time series that was created using the pixel-change method, which we detail next.

**Pixel change motion time series** Recall that the amount of pixel change within a video frame can be taken to index the amount of activity of a participant if they are the only source of movement in that part of the frame (Kupper et al., 2010; Paxton & Dale, 2013; Schmidt et al., 2012). To calculate the absolute difference of pixel change between adjacent video frames for both the child and the experimenter, we first split all of the video images recorded using the Kinect sensor down the middle into a child half and an experimenter half and then extracted image change time series from these separate video frame series (for examples of these time series see Fig. 3d and e). This was done by simply counting the number of pixels that changed color from one frame to the next. Because of the nature of the hand clapping game, in which the participant and experimenter crossed over their half of the frame repeatedly, it was impossible to obtain separate time series for each of them and therefore this task was excluded from the pixel change analysis.

### Data analyses

Prior to analysis all of the pre- and post-non-task relevant movement transient periods were cropped from the different



**Fig. 3** Sample movement time series from a child collected by the Kinect whole body movement in (a) the object tapping task, (b) the pointing task, and (c) the interpersonal hand clapping game. Also, sample movement

time series from the same child using the Pixel change video analysis in (d) the object tapping task and (e) the pointing task

time series. For comparison purposes all the final motion time series were then low-passed filtered using 10 Hz fourth order Butterworth filter to remove system measure noise. The same filter was used to investigate whether using the same analysis techniques applied to the Polhemus can simply be transported over to the Kinect or pixel change video analysis without having to create a whole new set of data reduction techniques.

To determine the stability and patterning of the social motor coordination that occurred between the children and the experimenter for each task and condition, one standard measure of interpersonal coordination was employed: distribution of relative phase (DRP) (see Schmidt & Richardson, 2008, for a review of studies that have employed this measure).

**Distribution of relative phase angles** This measure evaluated the concentration of relative phase angles between the movements of the child and experimenter (i.e., the relative space-time angular location of the movements of the child and experimenter) across nine  $20^\circ$  regions of relative phase ( $0\text{--}20^\circ$ ,  $21\text{--}40^\circ$ ,  $41\text{--}60^\circ$ ,  $61\text{--}80^\circ$ ,  $81\text{--}100^\circ$ ,  $101\text{--}120^\circ$ ,  $121\text{--}140^\circ$ ,  $141\text{--}160^\circ$ , and  $161\text{--}180^\circ$ ). To determine these distributions, we computed the continuous relative phase of the two time series between  $-180^\circ$  and  $180^\circ$  using the Hilbert transform (Pikovsky, Rosenblum, & Kurths, 2001). We then

computed the percentage of occurrence of the absolute value of the relative phase angles across the nine  $20^\circ$  regions of relative phase from  $0^\circ$  to  $180^\circ$ . Previous research has demonstrated that stable social motor coordination is characterized by a concentration of relative phase angles in the portions of the distribution near  $0^\circ$  and  $180^\circ$  (e.g., Fitzpatrick et al., 2013; Richardson, Marsh, Isenhower, Goodman, & Schmidt, 2007; Schmidt, Richardson, Arsenault, & Galantucci, 2007).

**Statistical analyses** DRP was analyzed using separate 9 (phase region:  $0^\circ$ ,  $30^\circ$ ,  $50^\circ$ ,  $70^\circ$ ,  $90^\circ$ ,  $110^\circ$ ,  $130^\circ$ ,  $150^\circ$ ,  $180^\circ$ )  $\times$  2 (diagnosis: ASD, TD) mixed ANOVAs for each task and motion-tracking method, with phase region as the repeated measures factor (a Greenhouse-Geisser correction was employed where necessary). Of particular interest was the difference in the magnitude of in-phase (percentage of time in  $0^\circ$  bin) and/or anti-phase (percentage of time in the  $180^\circ$  bin) for ASD and TD participants, and, therefore, when the mixed ANOVA phase region by diagnosis interaction was significant, planned *t*-tests were used to compare these two relative phase regions.

**Reliability analyses** In order to measure the overlap between the different measures, pairs of regressions were calculated

between the in-phase region occurrences of the Polhemus and every other motion-tracking system for each task.

## Results and discussion

### Object tapping task

All the results for the object tapping task ANOVAs for each of the motion capture methods can be found in Table 1.

**Wrist movement** A  $9 \times 2$  mixed ANOVA on the data obtained through the Polhemus system revealed a significant main effect of phase region and a significant phase region  $\times$  diagnosis interaction (see Table 1). Planned *t*-tests revealed that TD children had a significantly higher mean occurrence at 0 ( $M = 50.6, SD = 11.01$ ) than children with ASD ( $M = 37.92, SD = 15.38; t(36) = -2.92, p < .01, d = .95$ ; see Fig. 4a). As expected, both groups of children spent the majority of the trial in the 0 phase region, indicative of in-phase coordination.

There were no significant main effects nor an interaction for the DRP analysis of the Kinect forearm time series (see Table 1 and Fig. 4b). For this specific task the Kinect system did not have a good enough resolution to register the kinds of single limb, small-scale movements performed in the object tapping task used. Moreover, the drum-like cylinders used introduced sources of noise as can be seen when comparing the sample time series between the Polhemus and the Kinect data on Fig. 2a and d. This noise injection in turn led to the lack of differentiation between groups in DRP for the forearm data captured by the Microsoft Kinect.

**Whole body movement** The  $9 \times 2$  mixed ANOVA on the Kinect whole body vector movement time series DRP measures revealed a significant main effect of phase region (see Table 1). Planned *t*-tests showed that participants spent

significantly more time in the 180 phase region ( $M = 14.23, SD = 7.49$ ) than in the 0 phase region ( $M = 6.83, SD = 4.50; t(37) = -4.17, p < .01, d = 1.20$ ; see Fig. 4c). The phase region  $\times$  diagnosis interaction was not significant. No group differences were observed.

The  $9 \times 2$  mixed ANOVA on the pixel change DRP also revealed a significant main effect of phase region. As expected, both groups of children spent the majority of the trial in the 0 or in-phase region. Additionally, there was a significant phase region  $\times$  diagnosis interaction (see Table 1). Planned *t*-tests revealed that TD children had a significantly higher mean occurrence at 0 ( $M = 19.17, SD = 5.65$ ) than children with ASD ( $M = 5.22, SD = 15.38; t(36) = \times 2.17, p = .04, d = 1.20$ ; see Fig. 4d).

The Microsoft Kinect whole body measure therefore was not as accurate at capturing the movement dynamics of single limb, small scale movements such as those in this object tapping task. In particular, we fail to see differences between groups, and this method showed a significant difference in type of coordination that was contrary to what the Polhemus system showed. These results could be due to the level of noise in this situation created by the presence of the cylinders, which in turn led to skeletal tracking problems in parts of the task (compare Fig. 3a to Fig. 2a). This suggests that very careful planning is needed when designing the tasks being studied if the use of the Kinect skeletal tracker is being considered. It is vital to highlight again that the Kinect’s skeletal data is based on a combined infrared/video process of depth and a machine-learning algorithm that has a high noise to signal ratio. In this case, part of the motion being performed may have been confused with the cylinders in the skeletal tracker (this is true for both the forearm and the whole body methods).

The pixel change motion time series was, on the other hand, capable of capturing differences between the two groups in terms of the DRP measure as well as the type of coordination detected by the Polhemus system. In this case, the pixel change motion time series corresponded more closely to the Polhemus results presented above, than the time series extracted from the skeletal tracker data.

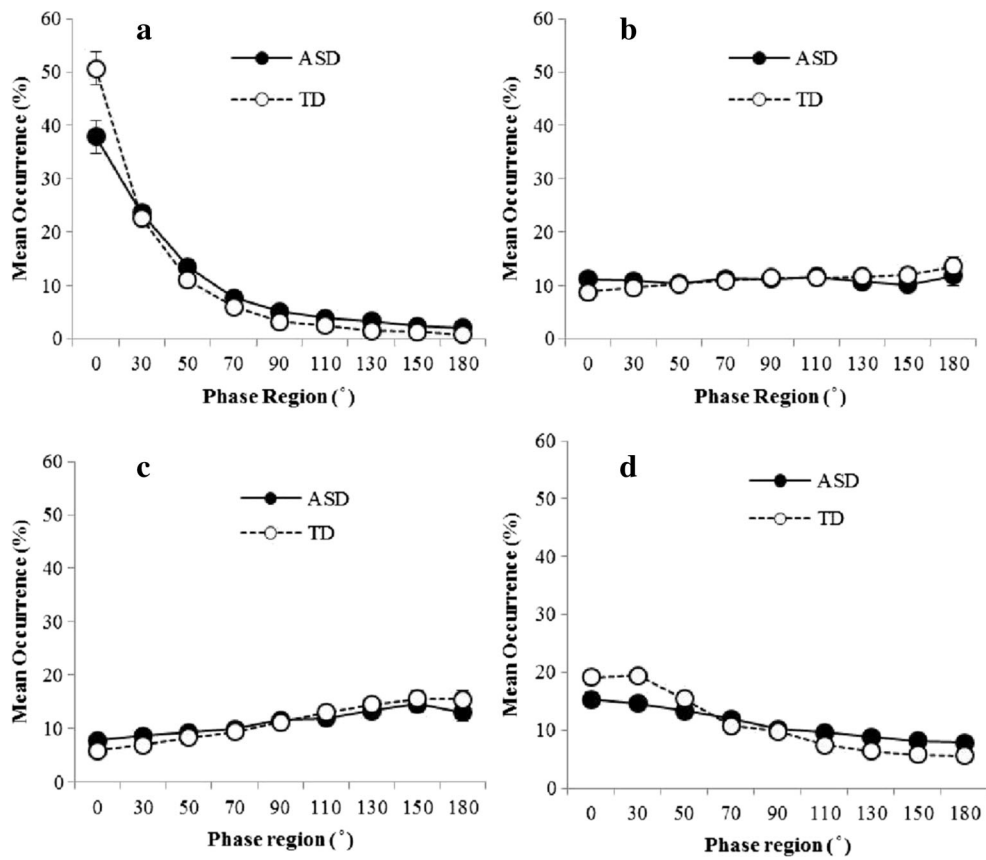
The reliability results corroborate the previous findings showing a negative correlation between the in-phase occurrence in the Polhemus as compared to the Kinect whole body ( $r = -.40$ ) and the Kinect forearm ( $r = -.19$ ) while showing a positive correlation with the video analysis ( $r = .27$ ; see Fig. 5). However, the regression between the video analysis in-phase occurrence and that measured by the Polhemus was not significant ( $p = .10$ ).

### Pointing task

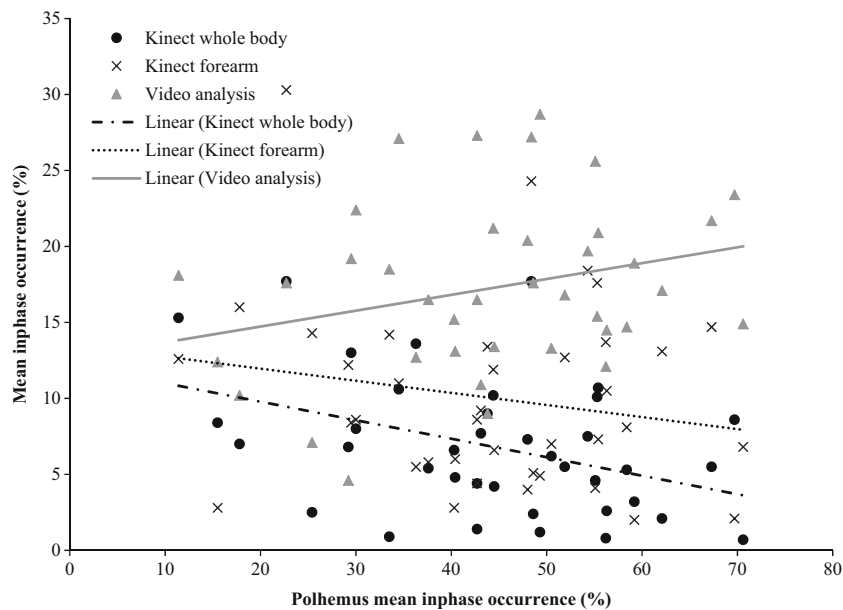
All the results for the pointing task ANOVAs for each of the motion capture methods can be found in Table 2.

**Table 1** Object tapping task ANOVA results

		MSE	<i>F</i>	<i>p</i>	$\eta^2_p$
Polhemus Latus	Phase region	32.49	237.36	< .01*	.87
	Diagnosis	0.001	0.75	.39	.02
	P $\times$ D	32.49	6.67	< .01*	.16
Kinect forearm	Phase region	26.55	0.95	.47	.03
	Diagnosis	0.001	0.0	1.00	0.0
	P $\times$ D	26.55	0.68	.71	.02
Kinect whole body	Phase region	29.22	11.76	< .01*	.25
	Diagnosis	0.001	0.11	.74	.003
	P $\times$ D	29.22	0.69	.70	.02
Pixel change video analysis	Phase region	13.00	51.30	< .01*	.59
	Diagnosis	0.001	1.70	.20	.05
	P $\times$ D	13.00	5.96	< .01*	.14



**Fig. 4** Distribution of relative phase in the object tapping task by group as measured by (a) the Polhemus Latus, (b) the Kinect: forearm movement, (c) the Kinect whole body vector, and (d) the video analysis. The error bars represent standard errors of the mean



**Fig. 5** Scatterplot showing the relationship between the mean in-phase percentage of occurrence in the Polhemus data in the tapping task against the Kinect whole body (represented by black circles and the linear trend

with a dotted and lined black line), the Kinect forearm (represented by Xs and the dotted gray line), and the video analysis (represented by the gray triangles and the solid gray line)



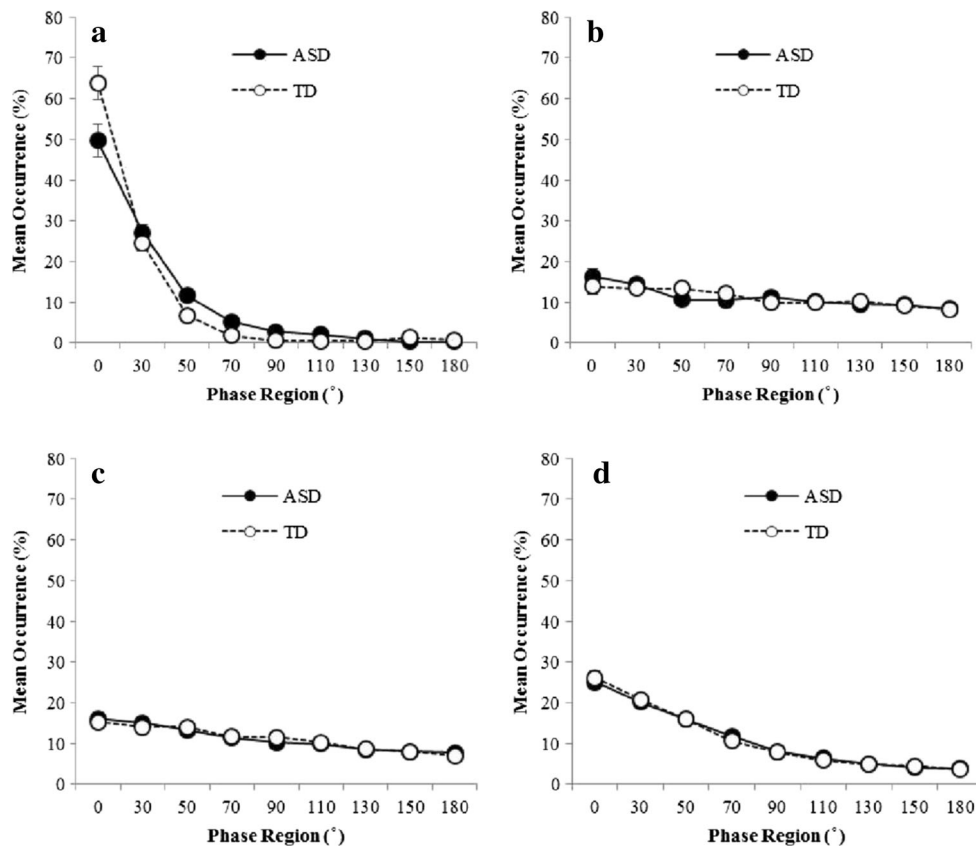
**Table 2** Pointing task ANOVA results

		MSE	<i>F</i>	<i>p</i>	$\eta^2_p$
Polhemus Latus	Phase region	58.04	235.43	< .01 *	.87
	Diagnosis	0.001	0.21	.65	.006
	<i>P</i> × <i>D</i>	58.04	5.14	.02 *	.13
Kinect forearm	Phase region	28.15	6.68	< .01 *	.16
	Diagnosis	0.001	1.50	.23	.04
	<i>P</i> × <i>D</i>	28.15	0.79	.44	.02
Kinect whole body	Phase region	18.64	18.02	< .01 *	.33
	Diagnosis	0.001	0.12	.73	.003
	<i>P</i> × <i>D</i>	18.64	0.29	.97	.01
Pixel change video analysis	Phase region	16.34	143.94	< .01 *	.80
	Diagnosis	0.001	0.41	.52	.01
	<i>P</i> × <i>D</i>	16.34	0.24	.81	.01

**Wrist movement** The analysis of the Polhemus Latus data's DRP data showed a significant main effect of phase region distribution and a significant phase region by diagnosis interaction (see Table 2). Planned *t*-tests showed that the mean occurrence of a 0 relative phase was significantly higher for the children in the TD group ( $M = 63.82$ ,  $SD = 18.88$ ) than those in the ASD group ( $M = 49.69$ ,  $SD = 16.51$ ;  $t(36) = -2.46$ ,  $p = .02$ ,  $d = .80$ ; see Fig. 6a).

The analysis performed on the Kinect forearm time series DRP data also revealed a significant main effect of phase region (see Table 2). Planned *t*-tests showed that participants spent significantly more time in the 0 phase region ( $M = 15.08$ ,  $SD = 8.22$ ) than the 180 phase region ( $M = 8.26$ ,  $SD = 4.79$ ;  $t(37) = 3.53$ ,  $p < .01$ ,  $d = 1.01$ ; see Fig. 6b). However, this method failed to capture the phase region × diagnosis shown with the Polhemus. No group differences were observed.

Compared to the previous tapping task, this tapping task involved more large-scale motion, which permitted us to register some of the differences in the Kinect skeletal forearm data. However, this method was still unable to differentiate between the groups. Again, it is vital to point out that the Kinect's skeletal data is based on a combined infrared/video process of depth and a machine-learning algorithm that has a high noise-to-signal ratio. In this case, part of the motion being performed was in front of the torso of the participants, and since the children sometimes wore long-sleeve shirts, which were the same color as their torso, this in turn led to potential skeletal tracking problems in parts of the task, (compare Fig. 2b to Fig. 2e). Once more, this suggests that very careful planning is needed when designing the tasks being studied to avoid occlusion when using the Kinect skeletal tracker is being considered.



**Fig. 6** Distribution of relative phase in the pointing task by group as measured by (a) the Polhemus Latus, (b) the Kinect: forearm movement, (c) the Kinect whole body vector, and (d) the video analysis. The error bars represent standard errors of the mean

**Whole body movements** The analysis of the whole body Kinect vector movement revealed a significant main effect of phase region (see Table 2). Planned *t*-tests showed that participants spent significantly more time in the 0 phase region ( $M = 15.66$ ,  $SD = 6.85$ ) than in the 180phase region ( $M = 7.33$ ,  $SD = 3.43$ ;  $t(37) = 5.63$ ,  $p < .01$ ,  $d = 1.54$ ; see Fig. 6c). The phase region  $\times$  diagnosis interaction was not significant. No group differences were observed.

The analysis of the pixel change data also showed a significant main effect of phase region (see Table 2). Planned *t*-tests showed that participants spent significantly more time in the 0 phase region ( $M = 25.57$ ,  $SD = 7.32$ ) than in the 180phase region ( $M = 3.71$ ,  $SD = 2.18$ ;  $t(37) = 15.48$ ,  $p < .01$ ,  $d = 4.05$ ; see Fig. 6d). The phase region  $\times$  diagnosis interaction was not significant. No group differences were observed.

With respect to the pointing task, therefore, all four measurement methods were able to index the intended relative phase coordination underlying the task. However, only the Polhemus motion-tracking system was powerful enough to differentiate the groups of children, indicating that for such single limb movements the low-cost Kinect and video analysis methods may not be appropriate for detecting subtle differences in coordination stability.

The reliability analysis for the pointing task showed a positive correlation between the Polhemus and the Kinect whole body ( $r = .14$ ), the Kinect forearm ( $r = .12$ ), and the video analysis ( $r = .44$ ; see Fig. 7). In this case, the video analysis was able to account for 19.3 % of the variance present in the Polhemus data ( $t = 2.93$ ;  $p = .01$ ). The other two regressions were not found to be significant (all  $p$ s  $> .40$ ). Therefore, if a

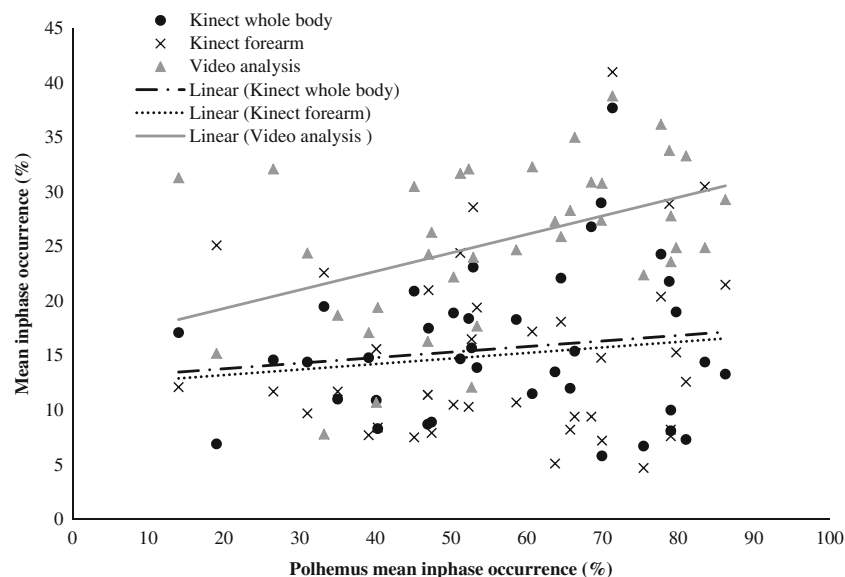
Polhemus system is not available to researchers in this type of task, a pixel change video analysis would be more reliable than using the Kinect skeletal tracker.

### Interpersonal hand clapping game

All the results for the interpersonal hand clapping game ANOVAs for each of the motion capture methods can be found in Table 3.

**Wrist movement** Analysis of the Polhemus Latus data showed a significant main effect of phase region and a significant phase region  $\times$  diagnosis interaction (see Table 3). Planned *t*-tests revealed a significantly lower occurrence for TD children in the 0 region ( $M = 0.06$ ,  $SD = 0.18$ ) than children in the ASD group ( $M = 0.96$ ,  $SD = 1.61$ ;  $t(35) = 2.36$ ,  $p = .02$ ,  $d = .79$ ). Additionally, the children in the TD group had a higher mean occurrence in the 180 phase region ( $M = 59.79$ ,  $SD = 13.03$ ) than those in the ASD group ( $M = 39.08$ ,  $SD = 18.12$ ;  $t(35) = -3.97$ ,  $p < .01$ ,  $d = 1.31$ ; see Fig. 8a).

Analysis of the Kinect forearm data also revealed a significant main effect of phase region and phase region by diagnosis interaction (see Table 3). Planned *t*-tests showed a significantly lower occurrence in the 0 region for TD children ( $M = 3.00$ ,  $SD = 2.91$ ) than children in the ASD group ( $M = 7.85$ ,  $SD = 5.31$ ;  $t(32) = 3.29$ ,  $p < .01$ ,  $d = 1.13$ ). Additionally, the children in the TD group had a higher mean occurrence in the 180phase region ( $M = 27.76$ ,  $SD = 12.16$ ) than those in the ASD group ( $M = 15.37$ ,  $SD = 9.32$ ;  $t(32) = -3.34$ ,  $p < .01$ ,  $d = 1.14$ ; see Fig. 8b).



**Fig. 7** Scatterplot showing the relationship between the mean in-phase percentage of occurrence in the Polhemus data in the pointing task against the Kinect whole body (represented by black circles and the linear trend

with a dotted and lined black line), the Kinect forearm (represented by Xs and the dotted gray line), and the video analysis (represented by the gray triangles and the solid gray line)

**Table 3** Interpersonal hand clapping game ANOVA results

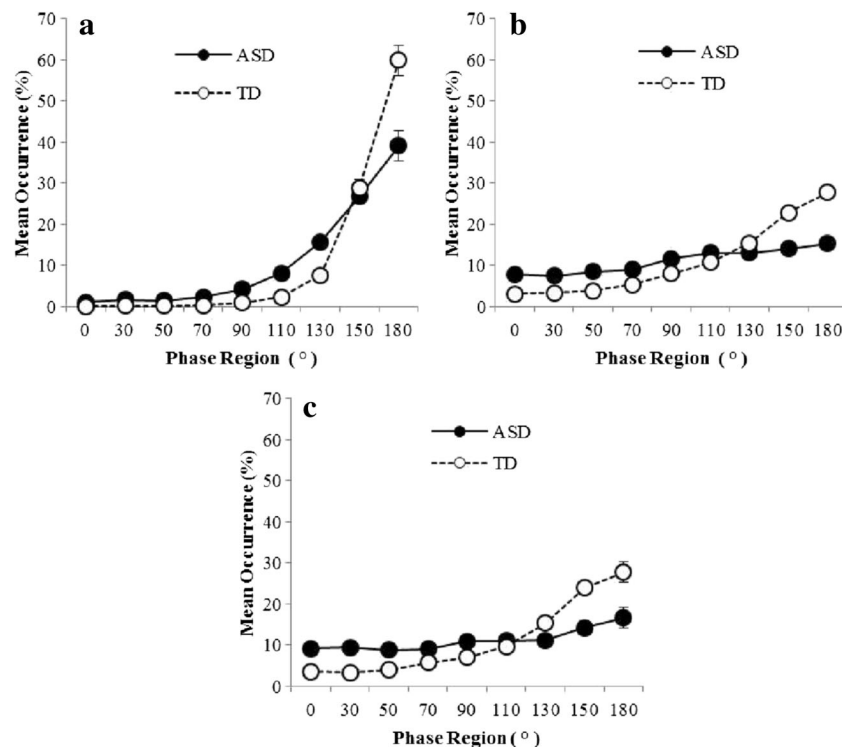
		MSE	<i>F</i>	<i>p</i>	$\eta^2_p$
Polhemus Latus	Phase region	49.88	210.92	< .01 *	.86
	Diagnosis	0.0005	2.27	.14	.06
	P × D	49.88	12.79	< .01 *	.27
Kinect forearm	Phase region	29.08	40.96	< .01 *	.56
	Diagnosis	0.001	2.22	.15	.07
	P × D	29.08	12.06	< .01 *	.27
Kinect whole body	Phase region	38.43	31.52	< .01 *	.49
	Diagnosis	0.001	0.04	.85	.001
	P × D	38.43	10.07	< .01 *	.23

**Whole body movement** Analyses performed on the Kinect whole body vector time series for the interpersonal hand clapping task revealed a significant main effect of phase region and a phase region by diagnosis interaction (see Table 3). Planned *t*-tests showed a significantly lower occurrence in the 0 region for TD children ( $M = 3.51$ ,  $SD = 4.57$ ) than children in the ASD group ( $M = 9.14$ ,  $SD = 7.22$ ;  $t(33) = 2.74$ ,  $p = .01$ ,  $d = .93$ ). Additionally, the children in the TD group had a higher mean occurrence in the 180 phase region ( $M = 27.72$ ,  $SD = 11.39$ ) than those in the ASD group ( $M = 16.61$ ,  $SD = 27.72$ ;  $t(33) = -3.16$ ,  $p < .01$ ,  $d = .52$ ; see Fig. 8c).

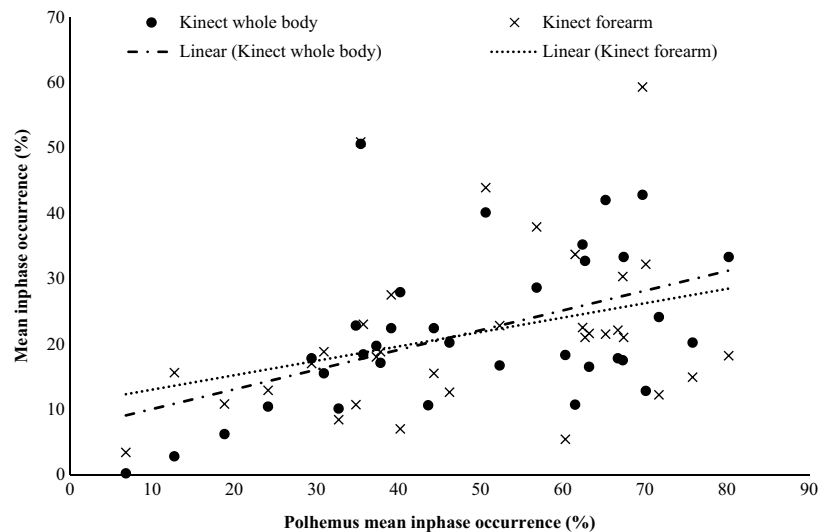
In summary, the Kinect forearm and whole body vector time series data were able to accurately capture the phase differences as well as differentiate the ASD and TD groups

for the interpersonal hand clapping game. Accordingly, this task seemed to be the most appropriate kind of task to be employed when using the Kinect as a motion capture tracking system. As can be seen when comparing Fig. 2c to Fig. 2f, the noise introduced in the Kinect tracking did not influence the shape of the time series as dramatically as it did with the previous two tasks. This task involved larger scale movements as well as movement of the torso and other body parts, while posing lower noise-inducing issues (since the hands and arms were not in front of the participants' torso, but to the side), leading to cleaner and more trustworthy results. This would indicate that the Kinect skeletal tracker system might be more suited to investigating movement coordination in tasks that involve the use of more gross, large-scale movements, especially those that involve multiple limbs. The pixel change method could not be used for this task because the hands of the two individuals contacted at midline, making it impossible to create separate time series of each individual. Additional research is needed to explore the reliability of the pixel change method in other gross, large-scale movements that do not involve body contact between the two individuals.

The reliability results in the hand clapping game confirmed the conclusions expressed above by showing a positive correlation between the in-phase occurrence between the Polhemus and the Kinect whole body ( $r = .50$ ) and the Kinect forearm ( $r = .34$ ; see Fig. 9). The Kinect forearm was able to capture 11.4 % of the variance present in the Polhemus in-phase occurrence ( $t = 2.00$ ;  $p = .055$ ), while the Kinect whole body was able to



**Fig. 8** Distribution of relative phase in the interpersonal hand clapping game by group as measured by (a) the Polhemus Latus, (b) the Kinect: forearm movement, and (c) the Kinect whole body vector. The error bars represent standard errors of the mean



**Fig. 9** Scatterplot showing the relationship between the mean in-phase percentage of occurrence in the Polhemus data in the hand clapping game against the Kinect whole body (represented by black circles and the linear

trend with a dotted and lined black line) and the Kinect forearm (represented by Xs and the dotted gray line)

account for 24.0 % of the variance present in the Polhemus data ( $t = 3.18$ ;  $p < .01$ ).

## General discussion

The goal of the current paper was to explicate whether low-cost motion-tracking systems could be used instead of a Polhemus Latus system when investigating social motor coordination in general and in children with ASD specifically. We used four methods of motion capture: (a) a high-end (yet moderately affordable) laboratory-grade Polhemus Latus magnetic motion-tracking system, (b) the Microsoft Kinect motion-tracking sensor, which is a low-cost optical tracking system, from which we extracted the limb movements as well as the (c) whole body movement, and (d) a video recording based pixel-change method of motion extraction. Of particular interest was how well the two low-cost Microsoft Kinect methods and the video pixel-change method performed in comparison to the more expensive, laboratory grade, Polhemus Latus system and the degree to which these three methods could be employed to differentiate the coordination that occurred in these tasks (in-phase, anti-phase) as well as the differences between the TD and ASD participants.

The findings demonstrated that the Polhemus Latus system did in fact provide a finer-grained measure of limb movement than the low-cost motion-tracking systems and was more robust in differentiating the groups both in measures of patterning and in stability of social coordination. This was the case across all three social motor tasks. This indicates, not surprisingly, that the Polhemus Latus system is superior for tasks that predominantly involve smaller-scale limb movements. However, the use of the Polhemus comes at a cost, for

example the wireless sensors needed in this method must be attached to the limbs in question, which can be problematic for certain participants (e.g., children with ASD). In addition, the system's reliance on magnetic signals makes its use incompatible with some other bio-behavioral measurement systems (e.g., EEG). Furthermore, when the number of limbs of interest increases above four, this method becomes cumbersome and more expensive for more complex, whole-body, naturalistic movements.

The use of the Kinect for measuring forearm movements was less successful than the Polhemus. In the tapping task not only did it fail to differentiate between the two groups, it also did not show the expected in-phase coordination pattern, and in the reliability analysis it also showed a negative correlation to the Polhemus. For the pointing task, the forearm skeletal tracker method was able to successfully measure the type of coordination expected in the task, but was unable to differentiate between the TD and ASD groups as measured by the Polhemus system. In this task, it showed the expected positive correlation to the Polhemus, but the regression failed to reach significance. However, this method was able to both measure the right type of coordination as well as differentiate the groups in the whole-body, naturalistic interpersonal hand clapping game and in the reliability analysis showed a positive correlation as well as a marginally significant regression with the Polhemus results. Furthermore, this method in cases involving special populations and children poses an advantage in future research because of its completely wireless set-up. That being said, special care is needed in designing tasks and setting up the data collection space to avoid occlusion if this tool is to be used.

In terms of the two methods measuring whole body movements, the current findings were mixed. The video pixel

change method was able to successfully measure the differences in coordination type expected in the pointing and tapping tasks and showed the expected positive correlations with the Polhemus results in the reliability analysis. Additionally, this method was able to detect differences between the two groups in the tapping task as well as a significant regression accounting for 19.3 % of the variance captured by the Polhemus, but failed to do so in the pointing task. The Kinect whole body method, on the other hand, found contradicting evidence regarding the type of coordination expected found by the Polhemus as well as the video pixel change methods in the object tapping task. It was, however, successful at differentiating the two coordination modes in the pointing task and interpersonal hand clapping game. Even though the whole body Kinect was unable to differentiate the groups in the first two tasks involving single limb movements, it was successful in the more naturalistic interpersonal hand clapping game, and even reached a significant level of prediction when regressed with the Polhemus in-phase results. As noted above, the major advantage of these methods is the freedom that researchers gain from completely wireless operation. While the video pixel change method would be simple to implement (since any video camera would be sufficient to use), it is worth noting once again that designing tasks needs special consideration since participants cannot cross over the middle of the frame at any point.

What is apparent at this point is that when employing these low-cost motion-tracking methods, particular care needs to be taken when designing the data collection space and the interaction tasks to be employed. In general the current results demonstrate that for both the Kinect sensor and pixel change methods, tasks with larger scale movements, as well as the use of more than just one or two limbs, provide the most accurate and reliable results. Of particular importance when using the Kinect is to choose tasks that have minimal occlusion issues (i.e., when the arms are not placed in front of the torso and when no props are used). When using the pixel change method the movements of the two people have to be in separate parts of the video frame and may be best suited to tasks involving less stereotyped movement as found, for example, in conversation. Since concluding the study, we have also found (through experimentation and also through literature) that if participants face the Kinect directly, the level of noise in the time series is reduced, therefore leading to better skeletal tracking. Additionally, Dutta (2012) found that by pronouncing the joints of interest simply by attaching “10cm wide circular discs or squares made of brightly coloured cardstock over a subject’s clothing” the skeletal tracking improved significantly. However, due to the reliance on the machine-learning algorithm built into the Kinect system, the results presented currently are preliminary. Accordingly, future work could record participants’ movements with the Kinect while recording their movement with Polhemus sensors (or other

motion capture systems) that correspond to the same skeletal markers in the Kinect in order to measure if the differences observed here are due to errors in the skeletal reconstruction or simple occlusion. Additionally, research is needed to investigate if the filtering methods used in the current paper are appropriate for the Kinect methods or if there are other more optimal solutions for the type of data recorded. Finally, it would be helpful to conduct a study in the future in which the data collection between the different methods is synchronized, such that each time-series can be directly compared to the others in order to fully understand how much each method’s capture coincides with the others.

In conclusion, the current study has shown that certain low-cost methods of motion tracking can be used to capture and index the coordination dynamics that occur between a child and an experimenter. More specifically, the results found the Polhemus system to be better than the Kinect and Video analysis methods, but that the latter two methods were still able to index motor coordination dynamics for tasks that involve larger scale limb and body movements. In other words, the Kinect cannot substitute a system like the Polhemus in small-scale, precision tasks such as the tapping and pointing task, but it can do so in a more whole-body naturalistic task such as the hand clapping game. As for the pixel change method, it cannot be used in tasks like the hand clapping game, but it can be suited to some smaller-scale tasks, such as tapping. More generally, the current study also validates previous research (e.g., Fitzpatrick et al., 2013) by demonstrating that children diagnosed with ASD show different social motor coordination patterns when compared to their typically developing counterparts. The low cost and completely wireless motion capture systems compared here can therefore provide researchers with new tools to explore social motor coordination and the role it plays not only in ASD, but also in other developmental delays disorders and social functioning pathologies (i.e., schizophrenia) as well as developing innovative treatment strategies with these tools (Chang, Chen, & Huang, 2011; Chung, Huang, Yeh, Chiang, & Tseng, 2014), as long as careful planning goes into designing the experimental tasks.

**Acknowledgments** The authors would like to thank Dr. Holly Barnard and Carrie Thomas for their help in data collection, and Caitlyn Coleman for her help in the pixel change analysis. Research reported in this publication was supported by the National Institutes of Health under Award Number R21MH094659. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## References

- Bernieri, F. J., & Rosenthal, R. (1991). Interpersonal coordination: Behavior matching and interactional synchrony. In R. B. Feldman & B. Rimé (Eds.), *Fundamentals of nonverbal behavior. Studies in*

- emotion & social interaction* (pp. 401–432). New York, NY: Cambridge University Press.
- Chang, Y., Chen, S., & Huang, J. (2011). A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities. *Research in Developmental Disabilities, 32*, 2566–2570.
- Chartrand, T., & Bargh, J. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology, 76*(6), 893–910.
- Chung, I., Huang, C., Yeh, S., Chiang, W., & Tseng, M. (2014). *Developing Kinect games integrated with virtual reality on activities of daily living for children with developmental delay*. In Y. Huang, H. Chao, D. Deng & J. H. Park (Eds.), (pp. 1091–1097) Springer Netherlands. doi:10.1007/978-94-007-7262-5\_124
- Dutta, T. (2012). Evaluation of the Kinect sensor for 3-D kinematic measurement in the workplace. *Applied Ergonomics, 43*(4), 645–649.
- Fitzpatrick, P., Diorio, R., Richardson, M. J., & Schmidt, R. C. (2013). Dynamical methods for evaluating the time-dependent unfolding of social coordination in children with autism. *Frontiers in Integrative Neuroscience, 7*, 1–13.
- Kinect for Windows Sensor Components and Specifications (2014). Retrieved April 21, 2014 from <http://msdn.microsoft.com/en-us/library/jj131033.aspx>
- Knoblich, G., & Sebanz, N. (2006). The social nature of perception and action. *Current Directions in Psychological Science, 15*, 99–104.
- Kupper, Z., Ramseyer, F., Hoffmann, H., Kalbermatten, S., & Tschacher, W. (2010). Video-based quantification of body movement during social interaction indicates the severity of negative symptoms in patients with schizophrenia. *Schizophrenia Research, 121*, 90–100.
- Liberty Latus Brochure (2012). Retrieved April 21, 2014 from [http://polhemus.com/\\_assets/img/LIBERTY\\_LATUS\\_brochure.pdf](http://polhemus.com/_assets/img/LIBERTY_LATUS_brochure.pdf)
- Marsh, K. L., Isenhower, R. W., Richardson, M. J., Helt, M., Verbalis, A. D., Schmidt, R. C., & Fein, D. (2013). Autism and social disconnection in interpersonal rocking. *Frontiers in Integrative Neuroscience, 7*(4), 1–8.
- Marsh, K. L., Richardson, M. J., & Schmidt, R. C. (2009). Social connection through joint action and interpersonal coordination. *Topics in Cognitive Science, 1*(2), 320–339. doi:10.1111/j.1756-8765.2009.01022.x
- Paxton, A., & Dale, R. (2013). Frame differencing methods for measuring bodily synchrony in conversation. *Behavior Research Methods, 45*, 329–334.
- Pikovsky, A., Rosenblum, M., & Kurths, J. (2001). *Synchrony: A universal concept in nonlinear sciences*. Cambridge: Cambridge University Press.
- Richardson, M. J., Marsh, K. L., Isenhower, R. W., Goodman, J. R., & Schmidt, R. C. (2007). Rocking together: Dynamics of intentional and unintentional interpersonal coordination. *Human Movement Science, 26*(6), 867–891.
- Schmidt, R. C., Richardson, M. J., Arsenuault, C. A., & Galantucci, B. (2007). Visual tracking and entrainment to an environmental rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 860–870.
- Schmidt, R. C., Morr, S., Fitzpatrick, P. A., & Richardson, M. J. (2012). Measuring the dynamics of interactional synchrony. *Journal of Nonverbal Behavior, 36*, 263–279.
- Schmidt, R. C., & Richardson, M. J. (2008). Dynamics of Interpersonal Coordination. In A. Fuchs & V. Jirsa (Eds.), *Coordination: Neural, Behavioral and Social Dynamics* (pp. 281–308). Heidelberg: Springer-Verlag.
- Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science, 1*(2), 353–367.
- Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., & Blake, A. (2011). Real-time human pose recognition in parts from single depth images. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1297–1304). IEEE Computer Society.