Proceedings of the 24th IEEE International
Symposium on Robot and Human Interactive Communication
Kobe, Japan, Aug 31 - Sept 4, 2015

Joint Action Perception to Enable Fluent Human-Robot Teamwork

Tariq Iqbal, Michael J. Gonzales, and Laurel D. Riek[1]

***Abstract*— To be effective team members, it is important for robots to understand the high-level behaviors of collocated humans. This is a challenging perceptual task when both the robots and people are in motion. In this paper, we describe an event-based model for multiple robots to automatically measure synchronous joint action of a group while both the robots and co-present humans are moving. We validated our model through an experiment where two people marched both synchronously and asynchronously, while being followed by two mobile robots. Our results suggest that our model accurately identifies synchronous motion, which can enable more adept human-robot collaboration.**

I. INTRODUCTION

In order for robots to competently and contingently collaborate with humans, they need to be able to solve the challenge of sensing high-level human activities occurring in their surroundings. This is particularly important when both the robots and people are moving [1], [2]. While recent advances in the field have improved robot perception in general (c.f., [3]), it is still difficult for a robot to recognize high-level human actions in the real-world, and use that information in a timely fashion to make informed decisions about their own actions [4].

For example, a robot might encounter people in a group performing various social actions, such as engaging in social games or synchronized movements. It can be difficult for a robot to perceive and understand all of these different types of events to make effective decisions. If a robot could make better sense of its environment, its interactions with humans would reach a higher level of coordination, resulting in a *fluent* meshing of actions [5]–[9].

Researchers from the activity recognition, robotics, and neuroscience communities generally consider psychomotor tasks to be comprised of *motor primitives* (also referred as motor schemas, control modules, and prototypes) [10]–[14]. The idea is that psychomotor tasks are comprised of "building blocks at different levels of the motor hiearchy" [12]. Stored primitives are syntactically combined to enable a wide range of complex actions. While there are theoretical debates in the aforementioned communities about the optimal way to model both primitives and higher-order tasks [15], [16], for practical purposes, researchers have successfully built working systems over the past few decades [17].

In our work, we focus on *kinematically-defined primitives*, which relate to sequences of movements made by the limbs in 3D space. Many successful approaches have been employed for detecting these primitives [17], with strong results realized for detecting a range of human behaviors from gross motor motion (walking, lifting) to manipulation (stacking objects, brushing teeth) [4], [18]–[23]. We explore the use of these primitives within the context of groups of humans and robots interacting.

The authors are from the Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN, 46556 USA. e-mail: {tiqbal, mgonza14, lriek}@nd.edu.

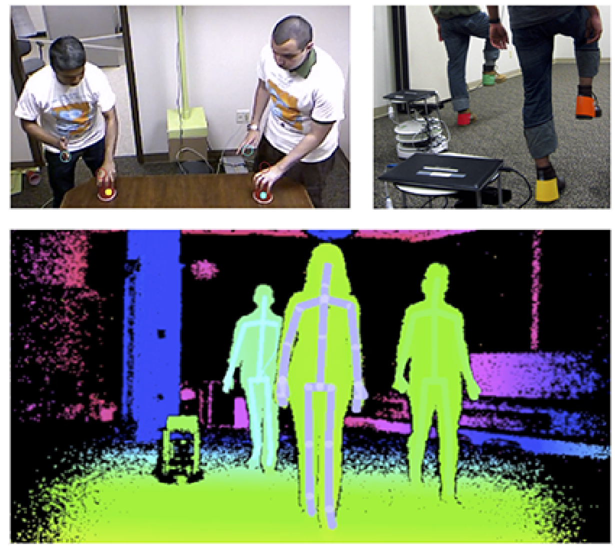

Fig. 1. Our research explores psychomotor entrainment between groups of people and robots, where we work toward enabling robots to automatically sense and respond to synchronous group behavior.

While this work is useful for some perceptual situations, from a human-robot teamwork perspective it may not tell robots much information about the *context* of how humans are interacting within the environment and with one another. This makes it challenging for robots to respond appropriately, and even more so when both the robots and co-present humans are moving [2], [24].

Synchronous motion, or joint action, is a common type of high-level behavior encountered in group interaction. This is a form of social interaction where two or more participants coordinate their actions in space and time while making changes to their environment [25]. Synchronous joint action is an important behavioral indicator of group-level cohesiveness, and also important for accurately understanding the affective behavior of a group [26], [27]. It is also a key component of human audio-visual attention and direction, and of social learning and cognitive development [28], [29]. Thus, there are many reasons why understanding synchronous joint action will help robots better perceive humans.

Researchers in the fields of cognitive science, kinesiology, and music have explored joint action measurement in human-human interaction (HHI), employing techniques such

978-1-4673-6703-5/15/$31.00 ©2015 IEEE

as modeling joint movement and human-behavior matching [26], [30]–[35]. Joint action has also recently surged in interest in the field of human-robot interaction (HRI) [36]–[46]. However, typically this research has been more geared toward dyadic HRI (one human, one robot), and has focused more on manipulation tasks.

In contrast, our work focuses on enabling robots to interact fluidly within groups of people, and focuses more on gross motion rather than dexterous manipulation [6], [7]. This is an important gap to address, as a large amount of human activity takes place in groups, and within groups there is a higher likelihood of synchronous activity occurring [26]. Thus, robots that can be aware of such activity and determine how best to engage are more likely to be accepted by humans [2].

If robots are to obtain this capability, however, they must first be able to accurately sense synchronous action occurring around them, ideally while in motion. To enable this, we present a method to automatically measure synchronous joint action between people as observed from mobile robots. Our method uses multiple types of task-level events performed by humans in the robot's environment to measure synchronous joint action. In this paper, we successfully validated our method against a rhythmic group activity (marching), as observed by two autonomous mobile robots (see Section III). In Section IV we present the results from this validation, and in Section V discuss their implications for the research community.

II. EVENT SYNCHRONIZATION MODEL

In an HHI scenario, each performer generates many interaction events, which result from the high-level tasks they perform. The timing as well as the outcome of each event depends on the events preceding it. The overall synchronization of a group depends on all of these task-level events.

Quian Quiroga et al. [47] proposed an event synchronization (ES) method which can be used for any time-series where events can be defined. This ES method is based on the relative timing of events. It is also possible to determine the leader-follower relationship between two time-series using the ES method, if one exists. Varni et al. [48] proposed an extension of this work to measure group synchronization using interaction patterns (i.e., synchronous motion). However, these methods only are able to incorporate homogeneous types of events (such as EEG signals or head motion trajectories).

To address this gap, we proposed a new method to automatically measure group psychomotor entrainment [49], [50]. In contrast to other techniques, our method takes multiple types of task-level events into account, is able to detect both asynchronous conditions, and is able to work with non-periodic time series data as it estimates entrainment.

We describe our method in detail in Section II-A and Section II-B. Here, we will first describe the method to measure synchronization of a human-robot group for single events. Later we will describe the extension of this model for multiple event types.

A. Measuring synchronization of a single type of event across two time-series

The task-level events associated with each individual involved in an interaction scenario can be expressed by a time-series. As described by Quian Quiroga et al. [47], suppose x_n and y_n are two time-series, where $n = 1 \ldots N$. Here, each time-series has N samples. Suppose, m_x and m_y are the number of events occuring in time-series x and y respectively, and E is the set of all events.

The events of both series are denoted by $e_x(i) \in E$ and $e_y(j) \in E$, where, $i = 1 \ldots m_x$, $j = 1 \ldots m_y$. The event times on both time-series are t_i^x and t_j^y ($i = 1 \ldots m_x$, $j = 1 \ldots m_y$) respectively. In the case of synchronous events in both time-series, the same event should appear roughly at the same time.

Two events are synchronous if the same event appears on both time-series within a time lag $\pm\tau$. Now, suppose $c^\tau(x|y)$ denotes the number of times a single type of event $e \in E$ appear in time-series x shortly after they appear in time-series y. Here,

$$c^\tau(x|y) = \sum_{i}^{m_x} \sum_{j}^{m_y} J_{ij}^\tau \tag{1}$$

where,

$$J_{ij}^\tau = \begin{cases} 1 & \text{if } 0 < t_i^x - t_j^y < \tau \\ \frac{1}{2} & \text{if } t_i^x = t_j^y \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Similarly, $c^\tau(y|x)$ denotes the number of times a single type of event $e \in E$ appear in time-series y shortly after they appear in time-series x. And,

$$c^\tau(y|x) = \sum_{j}^{m_y} \sum_{i}^{m_x} J_{ji}^\tau \tag{3}$$

where,

$$J_{ji}^\tau = \begin{cases} 1 & \text{if } 0 < t_j^y - t_i^x < \tau \\ \frac{1}{2} & \text{if } t_j^y = t_i^x \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$Q_\tau(e)$ represents the synchronization of events in two time-series, where we are only considering a single type of event e in both time-series. From $c^\tau(x|y)$ and $c^\tau(y|x)$, we can calculate $Q_\tau(e)$ as,

$$Q_\tau(e) = \frac{c^\tau(x|y) + c^\tau(y|x)}{\sqrt{m_x m_y}} \tag{5}$$

The value of $Q_\tau(e)$ should be in between 0 and 1 ($0 \leq Q_\tau(e) \leq 1$), as we normalize it by the number of events that appear in both time-series. $Q_\tau(e) = 1$ shows that all the events of both time-series are fully synchronized, and appeared within a time lag $\pm\tau$ on both time-series. On the other hand, $Q_\tau(e) = 0$ shows us that the events are asynchronous.

If there exists any leader-follower relationship of the events in time-series x and y, then $c^\tau(x|y)$ and $c^\tau(y|x)$ values give us that pattern [47]. This relationship can be incorporated during the calculation of synchrony for situations where this pattern might be important.

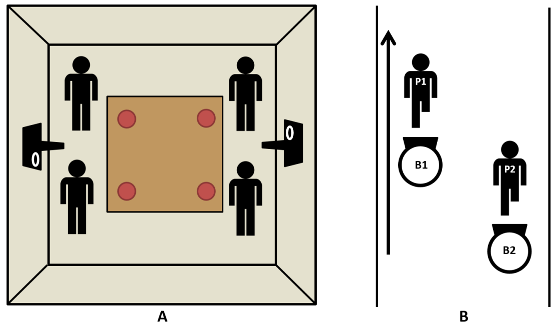

Fig. 2. A) First validation measured people performing synchronous joint action in a static setup [49]. B) Current validation method has autonomous mobile robots following humans moving synchronously. P1 and P2 are the performers, and B1 and B2 are the robots.

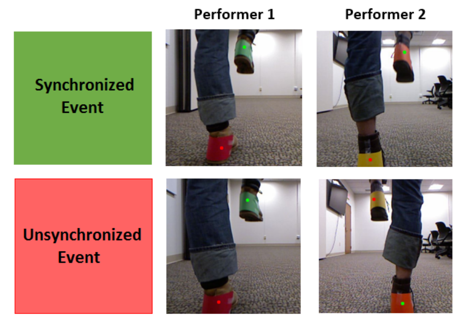

Fig. 3. A comparison of synchronous and asynchronous events captured by the mobile robots. Dots represent the tracked positions of the feet.

B. Measuring synchronization of multiple types of events across two time-series

When we are only considering a single type of event, $Q_\tau(e)$ gives us the synchronization of events in two time-series. Now, we extend the notion of synchronization of events in two time-series for multiple types of events.

Suppose we have n types of events $\{e_1, e_2, \ldots, e_n\} \in E(n)$, where $E(n)$ is the set of all types of events.. First, we calculate $Q_\tau(e_i)$ for each event type $e_i \in E(n)$. While calculating $Q_\tau(e_i)$, we will not consider any other type of event, except e_i. Now, let $m_x(e_i)$ be the number of events of type e_i occurring in time-series x and $m_y(e_i)$ is the number of events of type e_i occurring in time-series y.

To measure synchronization of multiple types of events between two time-series, we take the average of $Q_\tau(e_i)$ for each event type e_i, weighted by the number of events of that type. We will call this the synchronization index of that pair. So, the overall synchronization of events in time-series x and y of that pair is:

$$\forall e_i \in E(n) : Q_\tau^{xy} = \frac{\sum [Q_\tau(e_i) \times [m_x(e_i) + m_y(e_i)]]}{\sum [m_x(e_i) + m_y(e_i)]} \tag{6}$$

If all events are synchronous in both time-series, then the value of Q_τ^{xy} will be 1. On the other hand, when the events are not synchronous at all, the value of Q_τ^{xy} will be 0. We applied this model to a synchronized activity recorded using two mobile robots tracking two performers.

III. Method Validation

A. Method

In our prior work, we validated our method observing groups of people performing a synchronized psychomotor activity ("the cup game") using fixed sensors (see Fig. 2-left) [49], [50]. Here, we wanted to extend this work to explore how well our method worked with mobile sensors (robots) and mobile people. We were also interested in exploring asynchronous conditions. Thus, we analyzed the synchrony of people marching as followed by two mobile robots.

Marching is a group activity that is both dynamic as well as rhythmic in nature [51]. We conducted a set of controlled experiments where we had two individuals perform four different sets of marching actions (see Fig. 3). In each experiment, a mobile robot followed each performer to detect different events while the person was marching (see Fig. 2).

We used two Turtlebot robots as the mobile platforms in our experiment, and used their attached Kinect sensors for capturing data. The Turtlebot is an open-hardware and software platform comprised of an iRobot Create platform, a Microsoft Kinect, and an ASUS laptop running the Groovy version of the Robot Operating System (ROS) on Ubuntu Linux [52], [53].

Prior to our experiments, we adjusted the Kinect sensors on our Turtlebots so that they could track the performer's feet when they stepped and raised each foot. We used the TurtleBot's "Follower" program so that our robots could follow people autonomously while recording data, and stored the data in the ROS bag format.

We recruited two naive performers by word-of-mouth for our validation experiments. We gave both performers instructions for marching, and conducted one practice session before each of our four marching scenarios. We recorded two sessions for each of our four scenarios for analysis to account for noise.

In each scenario, we instructed each performer to perform a "high-march", picking up their knees and feet to an exaggerated degree. The first performer led the march, and was situated on the left side of the hallway. We situated the second performer behind and to the right of the first marcher as they moved down the hallway so that they could adjust their actions based on those of the first performer (see Fig. 2). All scenarios lasted approximately 35 seconds, and timed using a stopwatch. We gave the first performer an mp3 player with a set of noise-cancelling headphones playing John Philip Sousa's "Stars and Stripes Forever" march to keep tempo and prevent any noise-related distractions. Since the second performer is behind the first performer, the marching pace or pattern of the second performer did not affect the marching pace or pattern of the first performer.

In the first experimental scenario, the performers marched down a hallway in a synchronized manner. During this scenario, the second marcher followed the steps of the first marcher. In the second scenario, the second marcher followed

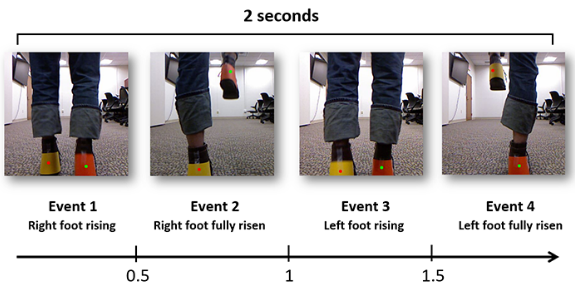

Fig. 4. A decomposition of events one through four that are used in event synchrony measurement. Overall, all events occur in two seconds.

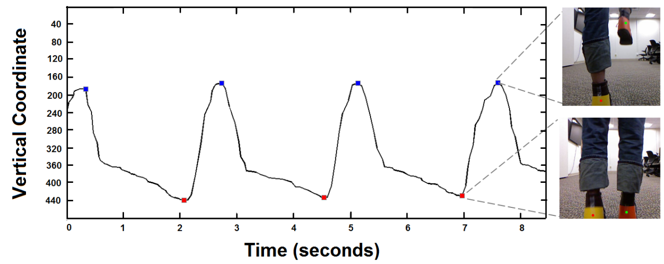

Fig. 5. This figure shows the detection of each event for one foot to show when a person's foot is leaving the ground, and when their foot reaches its maximum height.

the steps of the first marcher, but did so in the opposite order. For example, when the first marcher raised their left foot, the second marcher raised their right.

For our third scenario, both performers started out marching in a synchronous fashion, became asynchronous after 12 seconds, then synchronous again after 24 seconds. The fourth scenario reversed the actions of the third scenario, where both marchers began unsynchronized, became synchronized after 12 seconds, and unsynchronized after 24 seconds. We verbally told the second performer when to switch their steps to after these time periods by saying "switch". Fig. 3 shows the synchronous and asynchronous marching patterns.

B. Data Collection

Before recording data, both systems were synchronized with an Ubuntu time-server to ensure that both systems kept accurate timing. The videos of each performer were time-synchronized, and analyzed using our model to determine the group synchrony based on the occurrence of events.

To measure the overall group synchrony, we had to first determine the different task-level events of this group activity (i.e., a leg raise, or a leg leaving the ground) from the recorded videos. For this experimental setup, we defined two types of events to measure overall group synchrony. The first type of event was when a person begins to raise his/her a leg from the ground. The second type of event was when a leg reaches its maximum height. As a result, a total of four types of events occur when a person is marching (one of the aforementioned events for each leg). See Fig. 4 for an example of these events.

To track each performer's feet, we used a standard blob tracking technique. We used the ROS cmvision package to track color blobs in RGB images. We attached four unique small squares of colored paper (orange, yellow, green, and red) to the performers' left and right feet, while each robot followed behind each performer at a distance of two feet to track each performer's actions.

Fig. 6-A shows the expected synchrony for these four scenarios. We expected to see a high value for a synchronization index for the entire duration of a session for Scenario 1, and a value of zero for Scenario 2. For Scenario 3, we expected to see our measured synchronization index decrease beginning around seven seconds to a value of zero at 12 seconds, and increase again at about 20 seconds. For Scenario 4, we expected similar results, however in reverse order.

C. Data Analysis

From all detected blob positions, we first discarded very small blobs as noise. The remaining blob positions were considered candidates for a foot position. From these candidate blob positions, we generated an undirected graph by treating each blob center as a vertex of that graph. Two vertices were connected in this graph with an edge if they were closer to each other than a threshold distance.

From the resultant graph, we calculated the connected components. We then clustered all connected components together, and measured the total area of all blobs in each cluster. The cluster of blobs with the largest area was defined as the location of the foot in the RGB image. The center of that cluster, calculated by taking the mean position of each blob of that cluster weighted by each blob's area, is considered the foot position.

We then detected the high-level events in this rhythmic activity from the movements of each performer's feet. While marching, the foot position begins to move in an upward direction when the foot leaves the ground and eventually reaches its maxima before descending back downward [54], [55]. From the recorded video from the mobile robots, we found that this phenomenon also holds.

In our recorded video, the position of each foot changes significantly along the vertical axis of the RGB image plane while marching. Since the robot is also moving, we must extract the movements of the feet without regard to the robot's movement. Due to the ego-motion of the mobile robot and the attached camera, the position of the feet changed in every frame along both axes, although the feet were stationary in the real world. However, in the case of marching (raising the feet from the ground up), the change of the position of the feet along the vertical direction is significantly larger than the changes in position in the RGB image caused by the ego-motion.

To detect high-level events, we calculated the local extrema (maxima and minima) of the positions of the feet along the vertical axis. While calculating the local extrema, we use an additional condition that discarded any extrema that occurred within a specific period of time. This additional condition helps remove noise due to poor tracking or drastic movement caused by the robot's motion.

Due to the robot's ego-motion, the changes of the feet positions along the vertical axis were less than when they performed marching steps. An example of this effect can be seen between 3.2 and 4.6 seconds in Fig. 5. Our measure also accounts for this effect.

Our local minima and maxima are used to give us the time when a foot starts to leave the ground and when it reaches its maximum height. From these values, we are able to define that an event occurred during that time. This measure is independent to the ego-motion of the robot in our setup, as well as the height and pacing of each performer's movement. Fig. 5 describes this event detection model.

From the recorded video, we detected the events for both performers. We then measured the overall group synchrony using the model described in Section II.

IV. RESULTS

To sufficiently measure synchrony, we found a five second window to be ideal. This is due to the fact that each performer needed about one second to complete one step, and around two seconds for all four events to occur (one for each foot leaving the ground, and one for each foot reaching its maximum height, as shown in Fig. 4). We used $\tau = 0.21s$ and a sliding window of $5s$ for the calculation.

Each march session lasted about 35 seconds. Fig. 6 shows the synchronization index of each session over time. Here, four sessions are presented in four different graphs. As we used a sliding window size of five seconds, the synchronization index value for the zeroth second actually represents the synchronization index value on the window from zero to five seconds.

In Fig. 6, one can see that the synchronization index of the group is approximately 0.7 for the first experimental scenario, where the second performer was instructed to synchronously follow the steps of the first performer. In the case that the performers are asynchronous in their movements, then the synchronization index will be zero. On the other hand, if movement is very synchronous, then their synchronization index will be close to one.

The synchronization index of the second scenario is also presented in Fig. 6, which shows that the synchronization index is zero across the entire session. This indicates that during the time that our performers were marching, no event occurred in-sync over the entire session.

For the third experimental scenario, one can see that for the start of the session, the synchronization index was approximately 0.7 for the first few seconds, as shown in Fig. 6. This is due to the fact that we instructed performers to begin this session by marching in a synchronous manner. After 12 seconds, we verbally instructed the second performer to switch their steps and become asynchronous with the first performer, causing the synchronization index to decrease. After 24 seconds, we again verbally instructed the second performer to change their synchrony with the first performer using the term "switch". This caused an increase in the synchronization index.

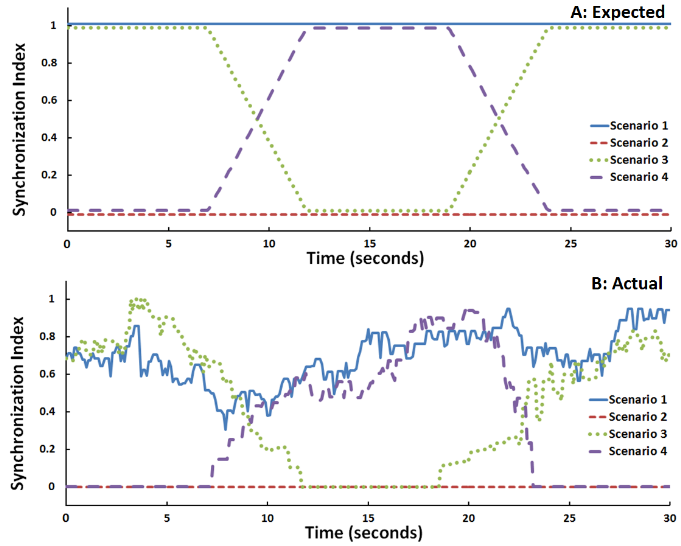

Fig. 6. A) The expected synchronization indices of our experimental scenarios. B) Actual synchronization indices of our experimental scenarios.

From Fig. 6, we can also see that the synchronization index became zero around the 12th second, which shows that the synchronization index for the five second window starting at the 12th second was zero. As we are using a sliding window of five seconds, the synchronization index began to decrease after approximately seven seconds. Similarly, when the second performer became synchronous with the first performer at the 24 second mark, we see a increasing synchronization index value starting at the 19th second mark.

For the fourth experimental scenario, one can see that the synchronization index was zero at the start of the session. This is because the performers began marching asynchronously until approximately the 12 second mark, where they became synchronous. After about the 24th second, marchers became asynchronous again. From the graph, we also see that the synchronization index is higher in the case of synchronous movements, which started to increase around the seventh second in the graph due to the sliding window algorithm. The synchronization index starts to decrease after the 20th second mark when movements are asynchronous after the 24th second, thereby dropping the synchronization index to zero.

We present the mean value of the synchronization indices of the sliding windows in Table I. In the first row, we present the average value of the synchronization indices of the sliding windows starting from zero to seven seconds. These sliding windows originally cover all of the events that occur between zero to 12 seconds. The second row shows the values of the sliding windows from the seventh to the 19th, and the last row presents from the 19th to the 30th second. The last two intervals represent the events occurring from the 12th to 24th second, and 24th to 35th second in real-time, respectively. We use these intervals because we instructed our second

TABLE I
MEAN SYNCHRONIZATION INDICES DURING EACH INTERVAL OF MARCHING

Time	Mean Synchronization Index			
	Scenario 1	Scenario 2	Scenario 3	Scenario 4
0-12s	0.67	0	0.8	0
12-24s	0.68	0	0.18	0.55
24-35s	0.79	0	0.68	0

marcher to switch the marching pattern at the 12th and the 24th second during the third and fourth scenarios.

From Table I one can see that the mean value of synchronization indices does not change much for Scenario 1 and 2, since the marching pattern did not change during these cases. For Scenario 1, the mean synchronization indices are high for all three time intervals. On the other hand, for all three intervals, we see the values are zero for Scenario 2. We see changes of the synchronization index value for Scenario 3 and 4 in different intervals. For Scenario 3, we observe lower values for the middle interval when the performers were asynchronous, and higher values for first and last interval when the performers were synchronous. We observe the opposite pattern for Scenario 4, with a higher value in the middle interval when the performers were synchronous, and lower in the first and the last interval where the performers were asynchronous.

V. DISCUSSION

As shown in Fig. 6, the synchronization indices for the four scenarios matched our expectations. The first scenario (synchronized) showed a high synchronization index (0.67) for the entire duration of the session. The second scenario (unsynchronized) similarly matched our expectations, with a value of zero for the entirety of the session. The third scenario (mixed, with a synchronized start) showed high values for time-frames that the second performer was instructed to be synchronized, and a value of zero for the time frame they were instructed to be unsynchronized. Scenario four (mixed, with an unsynchronized start) also matched these expectations.

Our results suggest that our method is effective in capturing and processing synchronized joint action occurring with both robots and people in motion. This work is encouraging for future work in understanding high-level group behavior detection and measurement in real-time for robotics. This work will also help in the design of robot behavior that generates synchronous joint action. Considering motion may distort sensing, our results show that our method was capable of detecting synchronized events and measuring synchronous joint action between two individuals in motion, independent of the height and pacing of steps.

While we expected a higher synchronization index value for Scenario 1, this could be due to a variety of reasons. Lighting and illumination variation may have an effect on blob tracking, especially since the robot and person are constantly in motion. Such changes may affect color calibration and tracking, causing the algorithm to miss a correct event or detect an incorrect event.

Our work is especially useful for human-robot collaboration. The results suggest that despite the difficulties in recognizing high-level group tasks for robots, we are able to detect synchronized events and measure synchronous joint action in motion using mobile robots. This work addresses the problem of detecting synchronized events that involve movements, despite current limitations that result from sensor noise and distortion due to motion.

In addition, our work may support researchers exploring human-robot fluency. Our method can enable a robot to have an automatic understanding of human motion in order to inform its own actions in response. This can play a role in handovers and collaborative manipulation [8], [9], [56].

This work is also useful for fields outside of robotics, such as social signal processing, kinesiology, and computer vision. Many researchers in these fields are interested in sensing and modeling human motion and using it to inform autonomous behavior; this work represents an important step that direction.

In terms of future work, this research represents the next step toward creating a real-time robotic system capable of understanding high-level group behavior to inform appropriate actions. Despite the constant motion of the performer, the constant movement of the robot, as well as variance in pacing in our study, our synchronization algorithm performed well in all scenarios. These findings inform our future work as we move to other scenarios, such as *enabling* joint action for robots interacting in groups, both online and in real-time. We have begun exploring this idea in human-robot group dance settings [57]. We also may explore alternative processing techniques to mitigate issues relating to light and color variation.

ACKNOWLEDGMENT

The authors thank Olivia Choudhury, Mark Costanzo, Maryam Moosaei, and Aaron Steele.

REFERENCES

[1] W. Choi, C. Pantofaru, and S. Savarese, "A general framework for tracking multiple people from a moving camera," *IEEE T Pattern Anal*, 2013.

[2] L. D. Riek, "The social co-robotics problem space: Six key challenges," in *Robotics: Science, and Systems (RSS), Robotics Challenges and Visions*, 2013.

[3] K. Buys, C. Cagniart, A. Baksheev, T. De Laet, J. De Schutter, and C. Pantofaru, "An adaptable system for rgb-d based human body detection and pose estimation," *J Vis Commun Image R*, 2013.

[4] H. Zhang and L. E. Parker, "4-Dimensional Local Spatio-Temporal Features for Human Activity Recognition," *IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, 2011.

[5] L. D. Riek, T. C. Rabinowitch, P. Bremner, A. G. Pipe, M. Fraser, and P. Robinson, "Cooperative gestures: Effective signaling for humanoid robots," in *Proc. of ACM/IEEE HRI*, 2010.

[6] T. Iqbal, M. J. Gonzales, and L. D. Riek, "A Model for Time-Synchronized Sensing and Motion to Support Human-Robot Fluency," in *Proc. of ACM/IEEE HRI, Workshop on Timing in HRI*, 2014.

[7] T. Iqbal, M. Gonzales, and L. D. Riek, "Mobile robots and marching humans: Measuring synchronous joint action while in motion," in *AAAI Fall Symposium Series*, 2014.

[8] G. Hoffman and C. Breazeal, "Cost-based anticipatory action selection for human–robot fluency," *IEEE Transactions on Robotics*, 2007.

[9] M. Cakmak, S. S. Srinivasa, M. K. Lee, S. Kiesler, and J. Forlizzi, "Using spatial and temporal contrast for fluent robot-human handovers," in *Proc. of ACM/IEEE HRI*, 2011.

[10] O. C. Jenkins, M. J. Matari, and S. Weber, "Primitive-Based Movement Classification for Humanoid Imitation," *IEEE-RAS Int'l Conference on Humanoid Robotics*, 2000.

[11] T. Flash and B. Hochner, "Motor primitives in vertebrates and invertebrates," *Current Opinion in Neurobiology*, 2005.

[12] G. Neumann, C. Daniel, and A. Kupcsik, "Information-Theoretic Motor Skill Learning," *AAAI 2013 Workshop on Intelligent Robotic System*, 2013.

[13] S. Schaal, A. Ijspeert, and A. Billard, "Computational approaches to motor learning by imitation," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 2003.

[14] R. A. Schmidt, "A schema theory of discrete motor skill learning." *Psychological review*, 1975.

[15] G. Salvendy, "Classification of human motions," *Theoretical Issues in Ergonomics Science*, 2004.

[16] A. Bobick and V. Krüger, "On human action," in *Visual Analysis of Humans*, 2011.

[17] J. Aggarwal and M. Ryoo, "Human activity analysis," *ACM Comput. Surv.*, 2011.

[18] K. Yun and J. Honorio, "Two-person interaction detection using body-pose features and multiple instance learning," *IEEE Comput. Vis. Pattern Recognit. Work. (CVPRW)*, 2012.

[19] B. Ni, G. Wang, and P. Moulin, "RGBD-HuDaAct: A Color-Depth Video Database for Human Daily Activity Recognition," *Consum. Depth Cameras Comput. Vis.*, 2013.

[20] H. Koppula and A. Saxena, "Learning spatio-temporal structure from rgb-d videos for human activity detection and anticipation," *ICML*, 2013.

[21] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy, "Berkeley MHAD: A comprehensive Multimodal Human Action Database," *2013 IEEE Work. Appl. Comput. Vis.*, 2013.

[22] S. Ali and M. Shah, "Human action recognition in videos using kinematic features and multiple instance learning." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010.

[23] M. J. Choi, A. Torralba, and A. S. Willsky, "Context Models and Out-of-context Objects," *Pattern Recognition Letters*, 2011.

[24] C. Hayes, M. O'Connor, and L. Riek, "Avoiding robot faux pas: Using social context to teach robots behavioral propriety," in *Proc. of ACM/IEEE HRI*, 2014.

[25] N. Sebanz, H. Bekkering, and G. Knoblich, "Joint action: bodies and minds moving together." *T. Cogn. Sci.*, 2006.

[26] M. J. Richardson, R. L. Garcia, T. D. Frank, M. Gergor, and K. L. Marsh, "Measuring group synchrony: a cluster-phase method for analyzing multivariate movement time-series," *Frontiers in Physiology*, 2012.

[27] G. Varni, G. Volpe, and A. Camurri, "A System for Real-Time Multimodal Analysis of Nonverbal Affective Social Interaction in User-Centric Media," *IEEE T Multimedia*, 2010.

[28] M. Rolf, M. Hanheide, and K. J. Rohlfing, "Attention via synchrony: Making use of multimodal cues in social learning," *IEEE T Autonomous Mental Development*, 2009.

[29] C. G. Prince, G. J. Hollich, N. A. Helder, E. J. Mislivec, A. Reddy, S. Salunke, and N. Memon, "Taking synchrony seriously: a perceptual-level model of infant synchrony detection," in *Int. Workshop Epigenetic Robot.*, 2004.

[30] L. Noy, E. Dekel, and U. Alon, "The mirror game as a paradigm for studying the dynamics of two people improvising motion together," *P Natl Acad Sci USA*, 2011.

[31] R. C. Schmidt and M. J. Richardson, "Dynamics of interpersonal coordination," *Und Com Sys*, 2008.

[32] S. Kirschner and M. Tomasello, "Joint drumming: social context facilitates synchronization in preschool children." *J Exp Child Psychol*, 2009.

[33] M. M. Louwerse, R. Dale, E. G. Bard, and P. Jeuniaux, "Behavior matching in multimodal communication is synchronized," *Cognitive Sci*, 2012.

[34] S. M. Boker and J. L. Rotondo, "Symmetry building and symmetry breaking in synchronized movement," *Adv Consc Res*, 2002.

[35] D. Lakens and M. Stel, "If They Move in Sync, They Must Feel in Sync: Movement Synchrony Leads to Attributions of Rapport and Entitativity," *Soc Cognition*, 2011.

[36] S. Lallée and et al., "Towards a platform-independent cooperative human-robot interaction system: Ii. perception, execution and imitation of goal directed actions," in *IROS 2011*, 2011.

[37] A. Mörtl, T. Lorenz, and S. Hirche, "Rhythm patterns interaction-synchronization behavior for human-robot joint action," *PloS one*, 2014.

[38] S. Nikolaidis, P. Lasota, G. Rossano, C. Martinez, T. Fuhlbrigge, and J. Shah, "Human-robot collaboration in manufacturing: Quantitative evaluation of predictable, convergent joint action," in *Int'l Symp. on Robotics (ISR)*, 2013.

[39] M. P. Michalowski, S. Sabanovic, and H. Kozima, "A dancing robot for rhythmic social interaction," in *Proc. of ACM/IEEE HRI*, 2007.

[40] K. Prepin and P. Gaussier, "How an agent can detect and use synchrony parameter of its own interaction with a human?" in *Dev. of Multimodal Interfaces: Active Listening and Synchrony*, 2010.

[41] S. K. Hasnain, G. Mostafaoui, and P. Gaussier, "A synchrony-based perspective for partner selection and attentional mechanism in human-robot interaction," *Paladyn*, 2013.

[42] G. Hoffman and G. Weinberg, "Gesture-based human-robot jazz improvisation," in *Proc. of IEEE ICRA*, 2010.

[43] A. D. Dragan, A. L. Thomaz, and S. S. Srinivasa, "Collaborative manipulation: new challenges for robotics and hri," in *Proc. of ACM/IEEE HRI*, 2013.

[44] L. Y. Morales Saiki, S. Satake, R. Huq, D. Glas, T. Kanda, and N. Hagita, "How do people walk side-by-side?: using a computational model of human behavior for a social robot," in *Proc. of ACM/IEEE HRI*, 2012.

[45] T. Kanda, H. Ishiguro, M. Imai, and T. Ono, "Body movement analysis of human-robot interaction," in *IJCAI*, 2003.

[46] B. Mutlu, A. Terrell, and C.-M. Huang, "Coordination mechanisms in human-robot collaboration," in *Proc. of ACM/IEEE HRI*, 2013.

[47] R. Q. Quiroga, T. Kreuz, and P. Grassberger, "Event synchronization: a simple and fast method to measure synchronicity and time delay patterns," *Phys Rev E*, 2002.

[48] G. Varni, G. Volpe, and B. Mazzarino, "Towards a social retrieval of music content," in *IEEE Conference on Social Computing*, 2011.

[49] T. Iqbal and L. D. Riek, "A Method for Automatic Detection of Psychomotor Entrainment," *Under Review*, 2014.

[50] T. Iqbal and L. Riek, "Assessing group synchrony during a rhythmic social activity: A systemic approach," in *6th Conference of the International Society for Gesture Studies (ISGS)*, 2014.

[51] J. H. Macdonald, "Lateral excitation of bridges by balancing pedestrians," *Proc. of the Royal Society A: Math., Phys. and Engg. Sci.*, 2009.

[52] "Robot Operating System," , Sept. 2014.

[53] L. D. Riek, "Embodied computation: An active-learning approach to mobile robotics education," *IEEE T Educ*, 2013.

[54] R. A. Mann and J. Hagy, "Biomechanics of walking, running, and sprinting," *The American journal of sports medicine*, 1980.

[55] A. Hreljac and R. N. Marshall, "Algorithms to determine event timing during normal walking using kinematic data," *J Biomech*, 2000.

[56] R. Wilcox, S. Nikolaidis, and J. A. Shah, "Optimization of temporal dynamics for adaptive human-robot interaction in assembly manufacturing." in *Robotics: Science and Systems*, 2012.

[57] S. Rack, T. Iqbal, and L. D. Riek, "Enabling synchronous joint action in human-robot teams," in *Proc. of ACM/IEEE HRI*, 2015.