Human-Robot Teaming: Approaches from Joint Action and Dynamical Systems

Tariq Iqbal and Laurel D. Riek

Contents

Abstract

As robots start to work alongside people, they are expected to coordinate fluently with humans in teams. Many researchers have explored the problems involved in building more interactive and cooperative robots. In this chapter, we discuss recent work and the main application areas in human-robot teaming. We also shed light on some practical challenges to achieving fluent human-robot

T. Iqbal (✉) • L.D. Riek
Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA, USA
e-mail: tiqbal@eng.ucsd.edu; lriek@eng.ucsd.edu

A. Goswami, P. Vadakkepat (eds.), *Humanoid Robotics: A Reference*,
https://doi.org/10.1007/978-94-007-7194-9_137-1

coordination and conclude the chapter with future directions for approaching these problems.

Keywords
Human-robot interaction · Human-robot teaming · Joint action · Dynamical group modeling · Coordination

1 Introduction

As robots are becoming more ubiquitous, they will be expected to interact with people in a range of settings, from dyads to groups. To be effective and functional teammates, robots need the ability to perceive and understand the activities performed by other group members. For example, if a robot can interpret various actions performed by people around it during a social event, then it can make efficient decisions about its own actions. However, it is difficult to automatically perceive and understand all the different tasks people engage in to make effective decisions as a teammate.

If a robot could make better sense of how humans interact among themselves in a group, its interactions with humans would reach a higher level of coordination, resulting in a fluent meshing of actions [13, 15, 25, 27, 29, 55, 62]. When two or more agents work together, Hoffman and Breazeal [15] defined fluency as the quality of achieving a high level of mutual coordination and adaptation. This quality is particularly important when the agents are well accustomed to the task and to each other.

This chapter discusses the existing methods and applications of human-robot interaction (HRI) in cooperative tasks. In many of these situations, robots are expected to work with people to achieve a common goal through the process of human-robot joint action. Thus, we start this chapter by giving a brief introduction to joint action, both in the context of human-human and human-robot joint action. We then summarize recent applications of human-robot cooperative interaction from the literature. Finally, we conclude the chapter by briefly presenting the challenges to realizing effective human-robot coordination with respect to hardware, software, and usability.

2 Background

2.1 Approaches from Cognitive Science to Model Joint Action

When a person acts alone, their behavior is very different than when they coordinate in a group [32]. When two or more persons coordinate in a group, it is important to understand the different ways they can interact among themselves and generate suitable interactive behaviors [30]. Many researchers from the fields of psychology and cognitive science investigate the underlying mechanisms of a joint action task. This includes how people interact together, how they understand the intention of other individuals, and how they coordinate together to perform a joint action. Curioni et al., chapter "▸ Joint Action in Humans: A Model for Human-Robot Interactions?" presented a detailed review of joint action in human teams.

Sebanz defined joint action as a form of social interaction where two or more participants coordinate their actions in space and time while making changes to their environment [33, 66]. Sebanz et al. described three important parts in a successful performance of a joint action task [65]. The first part makes a prediction about the intention of other interactional partners. The second involves understanding when to perform the actions jointly, as this is very important for temporal coordination. The last part involves understanding where and how to perform the joint action. The authors described these as the "what," "when," and "where" components of joint action.

Vesper et al. [76] suggested an architecture for joint action which focuses on planning, action monitoring, and action prediction processes and ways of simplifying coordination. This architecture described minimal requirements for an individual agent to engage in a joint action. This architecture aims to fill the gap between the approaches that focus on language and propositional attitudes and dynamical system approaches.

Many researchers have explored the underlying mechanisms that people may employ to perform a successful joint action task [44]. To perform joint actions successfully in a group, each individual needs to integrate self-behavior with a prediction about others' behavior simultaneously [48]. For example, Novembre et al. [48] investigated whether this integration process of self and other related behavior is underpinned by a neural process associated with motor simulation. They explored this through a music performance experiment. Their results suggested that motor simulation underpins temporal coordination during joint actions.

Other researchers took a group perspective approach to model a successful joint action. For example, Valdesolo et al. [74] investigated whether a coordinated action in a group has any influence on the ability of the group members to pursue a joint goal together. Their results suggested that a person's ability to rocking in synchrony enhanced that person's perceptual sensitivity to the motion of other group members. The ability to be synchronous with others resulted in an increase of their success in a joint action task.

Slowiński et al. [70] explored whether coordination between two people performing a joint action task is higher when they exhibit similar motion features. To explore this, they proposed an index of motion variability, called individual motor signature (IMS), to capture the subtle differences of human movements. They investigated the validity of this index via a mirror game. Their results suggested that when two people shared a similar IMS value, the synchronization level was higher.

2.2 Dynamical Modeling of Groups

In this subsection, we discuss the contrasting perspective, which is more bottom up and nonlinear, and explore coordination dynamics as a mechanism for realizing joint action. In group interactions, the activities of each member continually influence the activities of other group members. Most groups create a state of interdependence, where each member's outcomes and actions are determined in part by other members of the group [9]. This process of influence can result in coordinated group activity over time.

Many disciplines have approached the problem of how to assess coordination in a system. These include robotics, physics, neuroscience, psychology, dance, and music. Many of these techniques take a bottom-up approach, which first try to measure the low-level signals and then build a high-level behavior from the low-level signals [22, 23, 29, 41]. These low-level signals can include physical motion features, physiological features (e.g., heart rate), eye gaze behavior, or activity features. High-level behaviors, such as coordination within a group, are then inferred from these low-level signals.

For example, Richardson et al. [57] proposed a method to assess group synchrony by analyzing the phase synchronization of rocking chair movements. A group of six participants rocked in their chairs with their eyes either open or closed, and they used a cluster-phase method to quantify phase synchronization. Their results suggested that their group-level synchrony measure could successfully distinguish between synchronous and asynchronous conditions. Similarly, Néda et al. [43] investigated the development of synchronized clapping in a naturalistic environment. They quantitatively described the phenomena of how asynchronous group applause starts suddenly and transforms into synchronized clapping.

Coordination among explicit and implicit behaviors has also been explored in human-human interaction. Varni et al. [75] presented a system for real-time analysis of nonverbal, affective social interaction in a small group. In their study, several pairs of violin players performed while conveying four different emotions. The authors then used recurrence quantification analysis to measure the synchronization of the performers' affective behavior. In follow-on work, the researchers developed a system capable of analyzing the interaction patterns in a group of dancers.

Konvalinka et al. [35] explored coordination among implicit physiological signals and performed a study to measure the synchronous arousal between performers and observers during a Spanish fire-walking ritual. This synchronous arousal was derived from heart rate dynamics of the active participants and the audience.

Taking a nonlinear, dynamical systems approach, Iqbal and Riek developed a method to measure the degree of synchronous joint action in a group [20, 22, 24, 27, 29]. Their method takes multiple types of task level events into account while measuring the synchronization. This method can work on multiple types of heterogeneous events and can measure asynchronous situation in a group, in contrast to most other methods from the literature which only take a single type of event into account. The authors validated their method by applying it to both human-

human and human-robot teaming scenarios. Their results suggested that the method can successfully measure the degree of coordination in a group which matches the collective perception of group members. Extending this work, the authors designed a new approach to enable robots to perceive human group behavior in real time, anticipate future actions, and synthesize their own motion accordingly (see Fig. 1) [24, 29].

Lorenz et al. [38] also investigated movement coordination in human-human and human-robot teams. Their study involved both a human-human and human-robot dyad tapping on two positions on a table at certain times. The authors explored whether goal-directed, but unintentional, coordination of movements occurred during these interactions. Their results suggest that humans synchronized their movements with the movements of the robots.

3 Recent Applications

As robots are increasingly working with people, they need to perform joint actions with people efficiently. To achieve this, many of the aforementioned approaches have been employed in human-robot teams. This section will outline four main application areas where robots cooperatively perform joint action tasks with humans. We summarize the approaches used in these areas in Table 1.

3.1 Proximate Human-Robot Teaming

In many interactions, robots and humans need to share a common physical space to interact. Various methods are employed on robots to work efficiently in close proximities by avoiding collisions, such as models from a human demonstration, anticipatory action planning, etc. [72].

To build policies for robots to share a space with humans, many approaches in the literature first built models from human demonstrations. After training, robots then use these trained models to collaborate with people. For example, Ben Amor et al. [2] collected human motion trajectories as dynamic movement primitives (DMP) from a human-human task. After that, the authors used dynamic time warping to estimate the robot's DMP parameters. Using these parameters, they modeled human-robot joint physical activities using a new representation, called interaction primitives (IP). Their experimental results suggested that a robot successfully completed a joint physical task with a person when IPs were used.

Nikolaidis et al. [47] proposed a two-phase framework to fit a robot's collaborative policy to fit with a human collaborator. They first grouped the human activities into clusters and then learned a reward function for each cluster using an inverse reinforcement learning. This learned model was incorporated with a mixed observability Markov decision process (MOMDP) policy with the human type as the partially observable variable. After that, they used this model for a robot to infer the human type and to generate the appropriate policies.

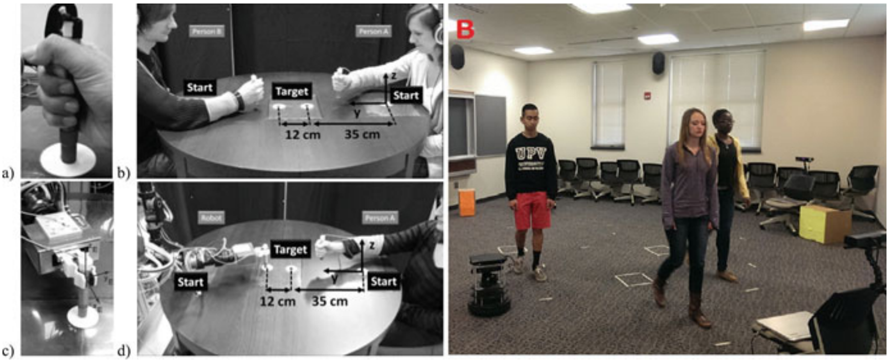

Fig. 1 People and robots are engaged in cooperative tasks (From [29, 38])

Many researchers try to achieve successful human-robot collaboration in a shared space by modeling human activities and use that knowledge as an input to a robot's anticipatory action planning mechanism [72]. This approach enables robots to generate movement strategies to efficiently collaborate with people.

For instance, Hoffman and Weinberg [17, 18] developed an autonomous robotic jazz-improvising robot, Simon, which played the marimba (see Fig. 2). To play in real time with a person, the robot needed an anticipatory action plan. The authors divided the actions into preparation and follow-through steps. Based on the anticipatory plans, their robot could simultaneously perform and react to shared activities with people.

Koppula et al. [36] also developed a method to anticipate a person's future actions. Anticipated actions were then used to plan appropriate actions for a robot to perform collaborative tasks in a shared environment. In their method, they model humans through low-level kinematics and high-level intent, as well as using contextual information. Then, they modeled the human's and robot's behavior through a Markov decision process (MDP). Their results suggested that this approach performed better than various baseline methods for collaborative planning.

Mainprice and Berenson [39] presented a framework to allow a human and a robot to perform a manipulation task together in close proximity. This framework used early prediction of the human motion to generate a prediction of human workspace occupancy. Then, they used a motion planner to generate robot trajectories by minimizing a penetration cost in the human workspace occupancy. They validated their framework via simulation of a human-robot collaboration scenario.

Along these lines, Pérez-D'Arpino et al. [52] proposed a data-driven approach which used human motions to predict a target during a reaching-motion task. Unhelkar et al. [73] extended this concept for a human-robot co-navigation task. This model used "human turn signals" during walking as anticipatory indicators of human motion. These indicators were then used to plan motion trajectories for a robot.

3.2 Human-Robot Handovers

A particular kind of activity often conducted in the proximate human-robot interaction space is a handover. It is an active application space in robotics research [72]. Most of the work on handovers focuses on designing algorithms for robots to successfully hand objects to people, as well as receive objects from them. The researchers working in this area use many methods to achieve their goals, including nonverbal signal analysis, human-human handover models, and legible trajectory analysis.

Many researchers used nonverbal signals of people to facilitate fluent object handover during human-robot interaction [72]. These signals included eye gaze, body pose, head orientation, etc. For example, Shi et al. [69] focused on building a model for a robot to handover leaflets in a public space, looking specifically at the relationship between gaze, arm extension, and approach. They used a pedestrian detector in their implementation on a small humanoid robot. Their results showed that pedestrians accepted more leaflets from the robot when their approach was employed than another state-of-the-art approach.

Similarly, Grigore et al. [10] demonstrated that the integration of an understanding of joint action into human-robot interaction can significantly improve the success rate of robot-to-human handover tasks. The authors introduced a higher-level cognitive layer which models human behavior in a handover situation. They particularly focused on the inclusion of eye gaze and head orientation into the robot's decision making.

Other researchers also investigated human-human handover scenarios to get inspiration to build models for human-robot handover scenarios [72]. Along this line of research, Huang et al. [19] analyzed data from human dyads performing a common household handover task – unloading a dish rack. They identified two coordination strategies that enabled givers to adapt to receivers' task demands, namely, proactive and reactive methods, and implemented these strategies on a robot to perform the same task in a human-robot team. Their results suggested that neither proactive nor reactive strategy can achieve both better team performance and better user experience. To address this challenge, they developed an adaptive method to achieve a better user experience with an improved team performance compared to the other methods.

To improve the fluency of a robot's actions during a handover task, Cakmak et al. [5] found that the failure to convey an intention of a robot to handover an object causes delay during the handover process. To address this challenge and to achieve fluency, the authors tested two separate approaches on a robot: performing distinct handover poses and performing unambiguous transitions between poses during the handover task. They performed an experiment where a robot used these two approaches while handing over an object to a person. Their findings suggested that unambiguous transition between poses reduced human waiting times, resulting in a smoother object handover. However, distinct handover poses did not have any effect on that.

Fig. 2 A live performance of a robotic marimba player (From [17])

Other researchers work on perform trajectory analysis to achieve smooth handover of objects. For example, Strabala et al. [71] proposed a coordination structure for human-robot handovers based on human-human handover. The authors first studied how people perform handovers with their partners. From this study, the authors structured how people approach, move their hands, and transfer objects. Taking inspiration from this structure, the authors then developed a similar handover structure for human-robot handover. This human-robot handover structure concerned about what, when, and where aspects of handovers. They experimentally validated this design structure.

3.3 Fluent Human-Robot Teaming

Many researchers in the robotics community try to build fluent human-robot teams. To achieve this goal, many approaches have been taken, including insights from human-human teams, cognitive modeling for robots, understanding the coordination dynamics of teams, and adaptive future prediction methods [72].

To achieve fluency in human-robot teams, many researchers investigated how people achieve fluent interaction in human-only teams. This knowledge is used to develop strategies for robots to achieve fluent interaction while interacting with people.

Taking insights from human-human teaming, Shah et al. [67, 68] developed a robot plan execution system, called Chaski, to use in human-robot teams. This system enables a robot to collaboratively execute a shared plan with a person. This system can schedule a robot's action and adapt to the human teammate to minimize the human's idle time. Through a human-robot teaming experiment, the authors validated that Chaski can reduce a person's idle time by 85%.

To build cognitive models for robots, researchers build on many other fields, including cognitive science, neuroscience, and psychology. For example, Hoffman and Breazeal [12] address the issue of planning and execution through a framework for collaborative activity in human-robot groups by building on the various notions from cognitive science and psychology literature. They presented a hierarchical goal-oriented task execution system. This system integrated human verbal and nonverbal actions, as well as robot nonverbal actions to support the shared activity requirements.

Iqbal, Rack, and Riek developed two anticipation algorithms for robots to coordinate their movements with people in teams by taking team coordination dynamics into account [29, 55]. One of the anticipation algorithms (SIA) relied on high-level group behavior understanding, whereas the other method (ECA) did not rely on high-level group behavior. The results indicated that the robot was more synchronous to the team and exhibited more contingent and fluent motion when the SIA method was used than the ECA method. These findings suggested that the robot performed better when it had an understanding of high-level group behavior than when it did not.

Additionally, Iqbal and Riek [24] investigated how the presence of robots affects group coordination when both their behavior and their number (single robot or multi-robot) vary. Their results indicate that group coordination is significantly affected when a robot joins a human-only group. The group coordination is further affected when a second robot joins the group and has a different behavior from the other robot. These results indicated that heterogeneous behavior of robots in a multi-human multi-robot group can play a major role in how group coordination dynamics stabilize.

Drawing inspiration from the neuroscience and cognitive science literature, Iqbal et al. [28] developed algorithms for robots which leveraged a humanlike understanding of temporal changes during the coordination process, with a particular eye toward an understanding of rhythmic tempo change. In their work, a robot employed two separate processes while coordinating with people, a temporal adaptation process, and a temporal anticipation process. A robot used the temporal adaptation process to compensate for temporal errors that occurred while coordinating with people. Additionally, the robot used the anticipation process to generate a prediction about the timing of the next action to coincide with the timing of the next external rhythmic signal. They applied these processes to a robot to drum synchronously with a group of people.

Building adaptive models based on a prediction of future actions is another approach to achieve fluent human-robot collaboration. Hoffman and Breazeal [15] developed a cognitive architecture for robots, taking inspiration from neuropsychological principles of anticipation and perceptual simulation. In this architecture, the fluency in joint action achieved through two processes: (1) anticipation based on a model of repetitive past events and (2) the modeling of the resulting anticipatory expectation as perceptual simulation. They implemented this architecture on a non-anthropomorphic robotic lamp, which performed a human-robot collaborative task. Their results suggested that the sense of team fluency and the robot's contribution

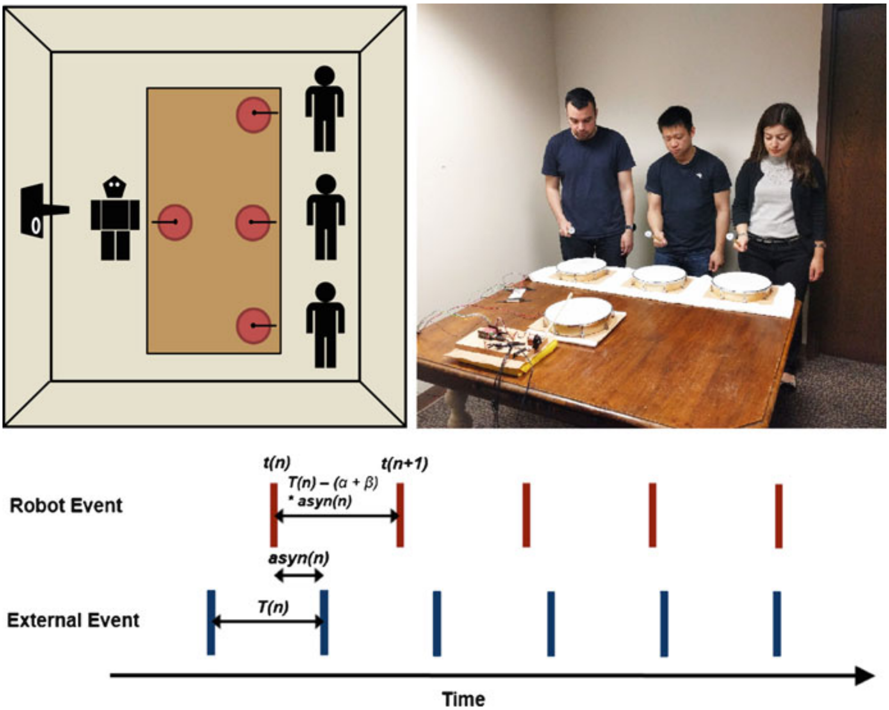

Fig. 3 A human-robot drumming team (From Iqbal et al. [28])

to the fluency significantly increased when the robotic lamp used their developed architecture.

In other work, Hoffman and Breazeal [14] proposed an adaptive action selection mechanism for a robot in the context of human-robot joint action. This model made anticipatory decisions based on the confidence of their validity and relative risk. They validated their model through a study involving human subjects working with a simulated robot. They used two versions of robotic behaviors during this study, one was fully reactive and another one used their proposed anticipation model. Their results suggested a significant improvement in best-case task efficiency and significant difference in the perceived commitment of the robot to the team and its contribution to the team's fluency and success.

3.4 Robot as a Partner

There are still many open areas regarding social interactional capabilities that a robot should have before it can fluently and naturally interact with people as a partner. Many researchers have tried to tackle these open questions by building models for robots to understand and to act appropriately as a partner in social situations (chapter "▶ Empathy as Signaling Feedback Between (Humanoid) Robots and Humans").

For example, Leite et al. [37] conducted an ethnographic study to investigate how a robot's capability of recognizing and responding empathically can influence an interaction. The authors performed the study in an elementary school where children interacted with a social robot. That robot had the capability of recognizing and responding empathically to some of the children's affective states. The results suggested that the robot's empathic behavior had a positive effect on how children perceived the robot (Fig. 3).

Many researchers also explored how a robot's explicit behavior can influence its interaction with people (chapters "▶ Enriching the Human-Robot Interaction Loop with Natural, Semantic, and Symbolic Gestures" and "▶ Movement-Based Communication for Humanoid-Human Interaction"). For example, Riek et al. [59, 61] investigated how imitation by a robot affects human-robot teaming. They designed a study where a robot performed three head gestures while interacting with a person: full head gesture mimicking, partial mimicking, and no mimicking. The authors found that in many cases, people nodded back in response to the robot's nodding during interactions. They suggested incorporating more gestures, along with head nods, while studying affective human-robot teaming.

In another study, Riek et al. [60] explored the effect of cooperative gestures performed by a humanoid robot in a teaming scenario. The authors performed an experiment where they manipulated the gesture type, the gesture style, and the gesture orientation performed by the robot while interacting with people. Their results suggested that people cooperate more quickly when the robot performed abrupt ("robotlike") gestures and when the robot performed front-oriented gestures. Moreover, the speed of people's ability to decode robot gestures is strongly correlated with their ability to decode human gestures.

In HRI, eye gaze can provide important nonverbal information [72]. For example, Moon et al. [40] performed an experiment where a robot performed humanlike gaze behavior during a handover task. In their experiment, a PR2 robot performed three different gaze behaviors while handing over a water bottle to a person. The results indicated that the timing of handover and the perceived quality of the handover event were improved when the robot showed a humanlike gaze behavior.

Admoni et al. [1] explored whether a deviation from a robot's standard behavior can influence the interaction. The authors claimed that people oftentimes overlooked robot's standard nonverbal signals (e.g., eye gaze) if they were not related to the primary task. In their experiment, the authors manipulated the handover behavior of a robot to deviate a little from the standard expected behavior. The results of this experiment suggested that a simple manipulation on standard handover timing of a robot made people be more aware of other nonverbal behaviors of the robot, such as eye gaze behavior.

Another well-investigated approach in the field is to teach a robot appropriate behaviors by teaching it through demonstration, i.e., learning from demonstration (LfD) [3]. For instance, Niekum et al. [45] developed a method to discover semantically grounded primitives during a demonstrated task. From these primi-

Table 1 Application areas of human-robot collaboration

Application areas	Approaches
Proximate human-robot teaming	Models from human demonstration ([2, 47]) Anticipatory action planning ([17, 18, 36, 39, 52])
Human-robot handovers	Nonverbal signal analysis ([10, 69]) Modeling based on human-human handover ([5, 19]) Trajectory analysis ([71])
Fluent human-robot teaming	Insights from human-human teams ([67, 68]) Cognitive modeling ([12, 24, 28, 29, 55]) Predicting actions ([14, 15])
Robot as a partner	Explicit behavior analysis ([37, 59, 60, 61]) Eye gaze analysis ([1, 40]) Learning from demonstration ([4, 11, 45])

tives, the authors then built a finite-state representation of the task. The authors used a beta-process autoregressive hidden Markov model to automatically segment demonstrations into motion categories. These categories were then further divided into motion grounded stated in a finite automaton. From many demonstrated examples, this model was trained on a robot.

Hayes [11] looked at mutual feedback as an implicit learning mechanism during an LfD scenario. The authors explored grounding sequences as a feedback channel for mutual understanding. In their study, both a person and a robot provided non-verbal feedback to communicate their mutual understanding. The results from the experiments showed that people provided implicit positive and negative feedback to the robot during the interaction, such as by smiling or by averting their gaze from the robot. The results of this work can help us to build adaptable robot policies in the future.

Brys et al. [4] explored how to merge reinforcement learning and LfD approaches together to achieve a better and faster learning phase. One key limitation of reinforcement learning is that it often requires a huge amount of training data to achieve a desirable level of performance. For a LfD approach, there is no guarantee about the quality of the demonstration, which can have many errors. Brys et al. investigated the intersection between these two approaches and tried to speed up the learning phase of RL methods using an approach called reward shaping.

4 Challenges

When a robot leaves controlled spaces and begins to work alongside people, many things taken for granted in terms of perception and action do not apply, because people act unpredictably, and little can be known about human environments in advance [46, 49, 58]. These Challenges, human-robot teamingchallenges include difficulties in human action detection, understanding of team dynamics, limitations in robot hardware and software design, and egocentric perception. This section introduces some of the challenges that researchers face while incorporating robots into human environments to coordinate with people and briefly discusses some solutions to these problems.

4.1 Uncertainty in Human Action Detection

One of the main challenges to detecting human actions is the unpredictability of human behavior. Sometimes it can be difficult for a robot to perceive and understand the different types of events involved in these activities to make effective decisions due to sensor occlusion, sensor fusion error, unanticipated motion, narrow field of view, cluttered backgrounds, etc. [6, 7, 56, 62].

One approach to address the challenge of human action detection is to use classification algorithms to detect actions from video data. However, this approach has major challenges, including intra- vs interclass variations between action classes, environment and recording settings, temporal variations of actions, and obtaining and labeling training data [53]. Moreover, using a classifier for action detection has several computational bottlenecks, including generalizability, abnormality detection, and classifier training [56].

Most of the approaches available in the literature cannot handle most of these challenges. Moreover, in most action recognition cases, researchers usually assume that camera positions are static. However, this is not the case for mobile robots [6].

Ryoo and Matthies try to address the challenge of action detection from a first-person point of view [63]. In their work, the authors try to detect seven classes of commonly observed activities during human-human interaction from a first-person point of view. Ryoo et al. [64] further extended this approach to detect early human activities from a robot. Using their method, a robot can detect human activities early, in real time, and in real-world environments. However, these methods still do not address other practical challenges, such as occlusion.

4.2 Unpredictable Changes in Team Dynamics

If a robot has some ability to model team dynamics, it can anticipate future actions in a team and adapt to those actions to be an effective teammate. However, understanding team dynamics is not trivial. If robots have an understanding of its environment, then its interactions within the team might facilitate a higher level of coordination.

In many human-human team situations, team members are explicitly assigned to various roles [50]. On the other hand, in many human-human teams, various roles emerge over time across the team members to achieve a common goal [34]. Oftentimes these assigned roles change dynamically based on necessities. For example, a person who begins to lead a team to move a table may follow another teammate's lead later during the moving process. How people coordinate and cooperate among themselves in these situations are important indicators for robots to understand various roles in groups.

In human-robot interaction scenarios, various role distribution models are used. High-level role distribution models in the HRI paradigm are master-slave, supervisor-subordinate, partner-partner, teacher-learner, and leader-follower [21, 26, 50] (chapter "▸ Applications in HHI Physical Cooperation"). However, these well-defined role distributions are rarely seen in real-world situations. Moreover, distributed roles change dynamically in many situations. Therefore, if the roles are not predefined for an interaction, the robot needs to make predictions about the role of copresent people, to infer its own role in the group.

Understanding the role of other people in a group is not easy for a robot. Thus, most of the human-robot teams are designed using some prior distribution of roles to achieve goals. However, a dynamic understanding of role distributions in a human-robot team can enable a robot to understand team dynamics more appropriately, which can lead to a fluent interaction in the group.

4.3 Limited Behavioral Versatility on Robots

Another challenge of incorporating robots in human teams is a lack of versatility of behaviors on robots. Most robots are designed to perform a specific task. Therefore, most of the time, they are limited in their behavioral abilities because they are restricted by their physical capabilities. For example, some robots are designed to perform manipulation tasks, some are good at recognizing and tracking people, and some are good at mobility.

However, a robot often needs to perform more than one of these abilities simultaneously to interact fluently and establish trust with people. For example, to socially interact with people, a robot needs to be able to identify them, approach them by avoiding obstacles, understand verbal and nonverbal messages, communicate verbally and nonverbally, and work alongside them. Thus, researchers need robots with versatile behaviors and abilities to build more efficient and functional human-robot teams.

Anthropomorphic robots are widely used in social environments to interact with people. These robots can engage with people in social interaction by perceiving various social cues from verbal and nonverbal channels and by communicating with people verbally and nonverbally. However, these types of robots are often not designed with capabilities to perform other tasks, such as mobility and manipulation. Kismet [31] was one of the first few anthropomorphic robots with an expressive face that was used to interact with people in social environments using gaze, facial expression, body posture, and vocal babbling. However, due to lack of other physical parts, such as hands, this social robot lacks the capability to perform hand gestures to interact with people fluently.

The Nao robot is a widely used humanoid robot for research [42], which can walk, show expressive gestures, and verbally communicate with people. Because of its expressive body gestures and verbal communication capabilities, it became a popular platform which enabled researchers to design a wide variety of interactions with people. However, it lacks facial expressions and is incapable of performing manipulation tasks, which limits its utility.

There are also non-anthropomorphic robots that interact and collaborate with people. These robots can show various verbal and nonverbal responses and can also generate animated gestures while collaborating with people. For example, Hoffman and Ju [16] designed a non-humanoid robot with expressive movements in mind. This robot can perform humanlike gestures, such as a head nod to express agreement and a head shake to express disagreement. This robot can express selective gestures; however, it cannot express many other gestures which are possible to perform with an expressive face.

On the other hand of the spectrum, there exist many robots that are strictly designed to perform manipulation tasks, e.g., Fetch and Freight robot by Fetch Robotics [8]. These arms are capable of performing dexterous manipulation tasks. However, these robots are not particularly functional in social situations, as oftentimes they are not safe around people, and cannot easily generate expressive behaviors.

PR2 robot by Willow Garage [54] is another widely used robotic platform, particularly for manipulation and handover research. This robot has two manipulation arms with grippers and can perform many dexterous tasks, which make it a widely used robotic manipulator by the research community. However, this robot lacks the capability to perform any expressive behavior toward people and not very suitable for human social environments.

4.4 Lack of Infrastructure to Support Replicability

Because of the wide range of platforms used on various robots, it is very challenging for researchers to replicate studies across different robots. This limitation prevents human-robot collaboration researchers from exploring the effects of using various kinds of robots in similar situations.

These difficulties include changes in sensor modalities across various platforms, variation in onboard processing units, and variation in physical structure. For example, if a robot has a high-definition RGB-D camera and has an onboard graphical processing unit, then it can detect facial expressions more precisely. On the other hand, if another robot only has a low-definition RGB camera with no onboard processing unit, then the same algorithms will not perform consistently.

The robot operating system (ROS) is a commonly used platform in the academic community [51]. However, as this is an open-source software, there are many challenges using it due to lack of software support and maintenance.

Moreover, similar algorithms need to be implemented on different platforms as not all robots are using a unified platform. This requires researchers to reimplement preexisting algorithms to accommodate different platforms, which oftentimes delay progress. Having common infrastructures will greatly help the research community to achieve replicability and to explore new robotic behaviors to coordinate with people.

5 Discussion

In this chapter, we discussed some exciting recent work on human-robot coordination. We briefly described recent approaches to model human-human and human-robot joint action from the literature. These approaches include neural process modeling, taking a group perspective, bottom-up approaches, nonlinear dynamical systems approaches, and implicit and explicit physiological signals.

We also discussed four main application areas in human-robot cooperation domain, namely, human-robot handovers, interaction in close physical proximities, fluent human-robot teaming, and robot as a partner. Many approaches have been taken to incorporate robots in these application domains, including dynamic trajectory analysis, anticipatory action planning, cognitive modeling, explicit and implicit behavior analysis, affective behavior analysis, and learning from demonstration (LfD).

Although there exist many applications of human-robot coordination, there also exist many practical issues that must be addressed to achieve a higher level of fluency in interaction. These practical issues include a lack of work done to detect and recognize copresent human actions and to understand team dynamics, limitations in robot design, and a lack of infrastructure to support replicability. Computational fields, like computer vision and machine learning, are trying to address specific robotic problems related to real-world scenarios, such as using egocentric vision, computationally inexpensive object proposal algorithms, and so on [6, 7]. Along with improvements in these technologies and existing algorithms, social robots will be able to cooperate with copresent people better in human social environments in the future.

6 Cross-References

▶ Applications in HHI Physical Cooperation
▶ Empathy as Signaling Feedback Between (Humanoid) Robots and Humans
▶ Enriching the Human-Robot Interaction Loop with Natural, Semantic, and Symbolic Gestures
▶ Joint Action in Humans: A Model for Human-Robot Interactions?
▶ Movement-Based Communication for Humanoid-Human Interaction

References

1. H. Admoni, A. Dragan, S. Srinivasa, B. Scassellati, Deliberate delays during robot-to-human handovers improve compliance with gaze communication, in *International Conference on Human-Robot Interaction*, Bielefeld (2014)
2. H. B. Amor, G. Neumann, S. Kamthe, O. Kroemer, J. Peters, Interaction primitives for human-robot cooperation tasks, in *IEEE Confernce on Robotics and Automation*, Hong Kong (2014)
3. B. Argall, S. Chernova, M. Veloso, A survey of robots learning from demonstration. Robot. Auton. Syst. **57**, 469–483 (2009)

4. T. Brys, A. Harutyunyan, V.U. Brussel, M.E. Taylor, Reinforcement Learning from Demonstration through Shaping, in *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, Buenos Aires (2015)
5. M. Cakmak, S.S. Srinivasa, M.K. Lee, S. Kiesler, J. Forlizzi, Using spatial and temporal contrast for fluent robot-human hand-overs, in *Proceedings of ACM/IEEE HRI*, Lausanne (2011)
6. D. Chan, L.D. Riek, Object proposal algorithms in the wild: are they generalizable to robot perception? in *Review* (2017)
7. D. Chan, A. Taylor, L.D. Riek, Faster robot perception using salient depth perception, in *IROS* (2017)
8. Fetch Robot (2017) https://www.fetchrobotics.com
9. D.R. Forsyth, *Group dynamics*, 4th edn. (T. Wadsworth, 2009)
10. E.C. Grigore, K. Eder, A.G. Pipe, C. Melhuish, U. Leonards, Joint action understanding improves robot-to-human object handover, in *IEEE/RSJ Internation Conference on Intelligent Robots and Systems (IROS)*, Tokyo (2013)
11. C.J. Hayes, M. Moosaei, L.D. Riek, Exploring implicit human responses to robot mistakes in a learning from demonstration task, in *Robot and Human Interactive Communication (RO-MAN)*, New York (2016)
12. G. Hoffman, C. Breazeal, Collaboration in human-robot teams, in *AIAA Intelligent Systems Technical Conference* (2007)
13. G. Hoffman, C. Breazeal, Cost-based anticipatory action selection for human–robot fluency. IEEE Trans. Robot. **23**(5), 952–961 (2007)
14. G. Hoffman, C. Breazeal, Effects of anticipatory action on human-robot teamwork efficiency, fluency, and perception of team, in *International Conference on Human-Robot Interaction*, Arlington (2007)
15. G. Hoffman, C. Breazeal, Anticipatory perceptual simulation for human-robot joint practice: theory and application study, in *AAAI*, Chicago (2008)
16. G. Hoffman, W. Ju, Designing robots with movement in mind. J. Hum.-Robot Interact **3**(1), 89–122 (2014)
17. G. Hoffman, G. Weinberg, Synchronization in human-robot Musicianship, in *International Symposium on Robot and Human Interactive Communication*, Viareggio (2010)
18. G. Hoffman, G. Weinberg, Interactive improvisation with a robotic marimba player in *Musical Robots and Interactive Multimodal Systems* (Springer, 2011), pp. 233–251
19. C.M. Huang, M. Cakmak, B. Mutlu, Adaptive coordination strategies for human-robot handovers, in *Robotics: Science and Systems* (IEEE Robotics and Automation Society, 2015)
20. T. Iqbal, L. Riek, Assessing group synchrony during a rhythmic social activity: a systemic approach, in *Proceedings of the Conference of the International Society for Gesture Studies (ISGS)*, San Diego (2014)
21. T. Iqbal, L.D. Riek, Role distribution in synchronous human-robot joint action, in *Proceedings of IEEE RO-MAN, Towards a Framework for Joint Action*, Portland (2014)
22. T. Iqbal, L.D. Riek, A method for automatic detection of psychomotor entrainment. IEEE Trans. Affect. Comput. **7**(1), 3–16 (2016)
23. T. Iqbal, L.D. Riek, Detecting and synthesizing synchronous joint action in human-robot teams, in: *International Conference on Multimodal Interaction*, Seattle (2015)
24. T. Iqbal, L.D. Riek, Coordination dynamics in multi-human multi-robot teams. IEEE Robot. Autom. Letters (RA-L) **2**(3), 1712–1717 (2017)
25. T. Iqbal, M.J. Gonzales, L.D. Riek, A model for time-synchronized sensing and motion to support human-robot fluency, in *ACM/IEEE Human-Robot Interaction, Workshop on Timing in HRI*, Bielefeld (2014)
26. T. Iqbal, M.J. Gonzales, L.D. Riek, Mobile robots and marching humans: measuring synchronous joint action while in motion, in *AAAI Fall Symposium on AI-HRI*, Arlington (2014)
27. T. Iqbal, M.J. Gonzales, L.D. Riek, Joint action perception to enable fluent human-robot teamwork, in *Proceedings of IEEE Robot and Human Interactive Communication*, Kobe (2015)

28. T. Iqbal, M. Moosaei, L.D. Riek, Tempo adaptation and anticipation methods for human-robot teams, in *Robotics: Science and Systems, Planning for HRI: Shared Autonomy and Collaborative Robotics Workshop*, Ann Arbor (2016).
29. T. Iqbal, S. Rack, L.D. Riek, Movement coordination in human-robot teams: a dynamical systems approach. IEEE Trans. Robot. **32**(4), 909–919 (2016)
30. N. Jarrassé, T. Charalambous, E. Burdet, A framework to describe, analyze and generate interactive motor behaviors. PLoS One **7**(11), e49945 (2012)
31. Kismet Robot, (2017), https://www.ai.mit.edu/projects/humanoid-robotics-group/kismet/kismet.html
32. G. Knoblich, J.S. Jordan, Action coordination in groups and individuals: learning anticipatory control. J. Exp. Psychol. Learn. Mem. Cogn. **29**(5), 1006–1016 (2003)
33. G. Knoblich, S. Butterfill, N. Sebanz, Psychological research on joint action: theory and data, in *Psychology of Learning and Motivation*. Advances in Research and Theory. **54** (2011), p. 59
34. I. Konvalinka, P. Vuust, A. Roepstorff, C.D. Frith, Follow you, follow me: continuous mutual prediction and adaptation in joint tapping. Q. J. Exp. Psychol. **63**(11), 2220–2230 (2010)
35. I. Konvalinka, D. Xygalatas, J. Bulbulia, U. Schjødt, E.M. Jegindø, S. Wallot, G. Van Orden, A. Roepstorff, Synchronized arousal between performers and related spectators in a fire-walking ritual. Proc. Natl. Acad. Sci. USA **108**(20), 8514–8519 (2011)
36. H.S. Koppula, A. Jain, A. Saxena, Anticipatory planning for human-robot teams. Springer Tracts Adv Robot **109**, 453–470 (2016)
37. I. Leite, G. Castellano, A. Pereira, C. Martinho, A. Paiva, Modelling empathic behaviour in a robotic game companion for children: an ethnographic study in real-world settings, in ACM/IEEE International Conference on Human-Robot Interaction, Boston (2012)
38. T. Lorenz, A. Mortl, B. Vlaskamp, A. Schubo, S. Hirche, Synchronization in a goal-directed task: human movement coordination with each other and robotic partners, in *Proc IEEE RO-MAN*, Atlanta (2011)
39. J. Mainprice, D. Berenson, Human-robot collaborative manipulation planning using early prediction of human motion, in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Tokyo (2013)
40. A. Moon, D.M. Troniak, B. Gleeson, M.K.X.J. Pan, M. Zheng, B.A. Blumer, K. MacLean, E.A. Croft, Meet me where I'm gazing: how shared attention gaze affects human-robot handover timing, in *ACM/IEEE International Conference Human-Robot Interaction*, Bielefeld (2014)
41. A. Mörtl, T. Lorenz, S. Hirche, Rhythm patterns interaction synchronization behavior for human-robot joint action. PloS One **9**(4), e95195 (2014)
42. Nao Robot, (2017), https://www.ald.softbankrobotics.com/en/cool-robots/nao
43. Z. Néda, E. Ravasz, Y. Brechet, T. Vicsek, A.L. Barabási, Self-organizing processes: the sound of many hands clapping. Nature **403**(6772), 849–850 (2000)
44. R.D. Newman-Norlund, M.L. Noordzij, R.G.J. Meulenbroek, H. Bekkering, Exploring the brain basis of joint action: co-ordination of actions, goals and intentions. Soc Neurosci **2**(1), 48–65 (2007)
45. S. Niekum, S. Chitta, Incremental semantically grounded learning from demonstration. Robot. Sci. Syst. **IX** (2013)
46. A. Nigam, L.D. Riek, Social context perception for mobile robots, in *IEEE/RSJ Intelligent Robots and Systems (IROS)*, Hamburg (2015)
47. S. Nikolaidis, K. Gu, R. Ramakrishnan, J. Shah, R.O. May, Efficient model learning for human-robot collaborative tasks (2014), pp. 1–9. https://doi.org/10.1145/2696454.2696455
48. G. Novembre, L.F. Ticini, S. Schütz-Bosbach, P.E. Keller, Motor simulation and the coordination of self and other in real-time joint action. Soc. Cogn. Affect. Neurosci. **9**(8), 1062–1068 (2014)
49. M.F. O'Connor, L.D. Riek, Detecting social context: a method for social event classification using naturalistic multimodal data, in *Automatic Face and Gesture Recognition (FG)*, Ljubljana (2015)
50. K. Ong, G. Seet, S. Sim, An implementation of seamless human-robot interaction for telerobotics. Int. J. Adv. Robot. Syst. **5**(2), 18 (2008)

51. Open Source Robotics Foundation, (2017), https://www.osrfoundation.org/
52. C. Pérez-DArpino, J. Shah, Fast target prediction of human reaching motion for cooperative human-robot manipulation tasks using time series classification, in *International Conference on Robotics and Automation*, Seattle (2015)
53. R. Poppe, A survey on vision-based human action recognition. Image Vis. Comput. **28**(6), 976–990 (2010)
54. PR2 Robot, (2017), http://www.willowgarage.com/pages/pr2/overview
55. S. Rack, T. Iqbal, L. Riek, Enabling synchronous joint action in human-robot teams, in *Proceedings of ACM/IEEE Human-Robot Interaction*, Portland (2015)
56. M. Ramanathan, Y. Wy, E.K. Teoh, Human action recognition with video data : research and evaluation challenges. IEEE Trans. Hum.-Mach. Syst. **44**(5), 650–663 (2014)
57. M.J. Richardson, R.L. Garcia, T.D. Frank, M. Gergor, K.L. Marsh, Measuring group synchrony: a cluster-phase method for analyzing multivariate movement time-series. Front. Physiol. **3**, 405 (2012)
58. L.D. Riek, The social co-robotics problem space: six key challenges, in Proceedings of RSS, Robotics Challenges and Visions (2013)
59. L.D. Riek, P. Robinson, Real-time empathy: Facial mimicry on a robot, in *International Conference on Multimodal Interfaces, Affective Interaction in Natural Environments (AFFINE)*, Chania (2008)
60. L.D. Riek, T.C. Rabinowitch, P. Bremner, A. Pipe, M. Fraser, P. Robinson, Cooperative gestures: effective signaling for humanoid robots, in *ACM/IEEE International Conference on Human-Robot Interaction*, Osaka (2010)
61. L.D. Riek, P.C. Paul, P. Robinson, When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry. J. Multimodal User Interfaces **3**, 99–108 (2010)
62. L.D. Riek, T.C. Rabinowitch, P. Bremner, A.G. Pipe, M. Fraser, P. Robinson, Cooperative gestures: effective signaling for humanoid robots, in *Proceedings of ACM/IEEE HRI*, Osaka (2010)
63. M.S. Ryoo, L. Matthies, First-person activity recognition: what are they doing to me? in *Proceedings of IEEE Computer Vision and Pattern Recognition*, Portland (2013)
64. M.S. Ryoo, T.J. Fuchs, L. Xia, J. Aggarwal, L. Matthies, Robot-centric activity prediction from first-person videos: what will they do to me? in *Proceedings of ACM/IEEE HRI*, Portland (2015)
65. N. Sebanz, G. Knoblich, Prediction in joint action: what, when, and where. Top. Cogn. Sci. **1**(2), 353–367 (2009)
66. N. Sebanz, H. Bekkering, G. Knoblich, Joint action: bodies and minds moving together. Trends Cogn. Sci. **10**(2), 70–76 (2006)
67. J. Shah, C. Breazeal, An Empirical Analysis of Team Coordination Behaviors and Action Planning With Application to Human-Robot Teaming. Hum. Factors J. Hum. Factors Ergon. Soc. **52**(2), 234 (2010)
68. J. Shah, J. Wiken, B. Williams, C. Breazeal, Improved human-robot team performance using Chaski, a human-inspired plan execution system, in *Proceedings of 6th International Conference on Human-Robot Interaction*, Lausanne (2011)
69. C. Shi, M. Shiomi, C. Smith, T. Kanda, H. Ishiguro A model of distributional handing interaction for a mobile robot in *Robotics: Science and Systems* (2013), pp. 24–28
70. P. Słowiński, C. Zhai, F. Alderisio, R. Salesse, M. Gueugnon, L. Marin, B.G. Bardy, M. di Bernardo, K. Tsaneva-Atanasova, Dynamic similarity promotes interpersonal coordination in joint-action. J. R. Soc. Interface **13**(116), 20151093 (2016)
71. K. Strabala, M.K. Lee, A. Dragan, J. Forlizzi, S.S. Srinavasa, M. Cakmak, V. Micelli, Towards seamless human-robot handovers. J. Hum.-Robot Interact. **2**(1), 112–132 (2013)
72. A. Thomaz, G. Hoffman, M. Cakmak, Computational human-robot interaction. Found. Trends Robot. 4(2–3), 105–223 (2016)
73. V.V. Unhelkar, C. Pérez-DArpino, L. Stirling, J. Shah, Human-robot co-navigation using anticipatory indicators of human walking motion, in *IEEE Conference on Robotics and Automation*, Seattle (2015)

74. P. Valdesolo, J. Ouyang, D. DeSteno, The rhythm of joint action: synchrony promotes cooperative ability. J. Exp. Soc. Psychol. **46**(4), 693–695 (2010)
75. G. Varni, G. Volpe, A. Camurri, A system for real-time multimodal analysis of nonverbal affective social interaction in user-centric media. IEEE Trans. Multimedia **12**(6), 576–590 (2010)
76. C. Vesper, S. Butterfill, G. Knoblich, N. Sebanz, A minimal architecture for joint action. Neural Netw. **23**(8–9), 998–1003 (2010)