Recognition of Haptic Interaction Patterns in Dyadic Joint Object Manipulation

Cigil Ece Madan, Ayse Kucukyilmaz, Tevfik Metin Sezgin, Cagatay Basdogan

Recognition of Haptic Interaction Patterns in Dyadic Joint Object Manipulation

Cigil Ece Madan, Ayse Kucukyilmaz, Tevfik Metin Sezgin, and Cagatay Basdogan

Abstract—The development of robots that can physically cooperate with humans has attained interest in the last decades. Obviously, this effort requires a deep understanding of the intrinsic properties of interaction. Up to now, many researchers have focused on inferring human intents in terms of intermediate or terminal goals in physical tasks. On the other hand, working side by side with people, an autonomous robot additionally needs to come up with in-depth information about underlying haptic interaction patterns that are typically encountered during human-human cooperation. However, to our knowledge, no study has yet focused on characterizing such detailed information. In this sense, this work is pioneering as an effort to gain deeper understanding of interaction patterns involving two or more humans in a physical task. We present a labeled human-human-interaction dataset, which captures the interaction of two humans, who collaboratively transport an object in an haptics-enabled virtual environment. In the light of information gained by studying this dataset, we propose that the actions of cooperating partners can be examined under three interaction types: In any cooperative task, the interacting humans either 1) work in harmony, 2) cope with conflicts, or 3) remain passive during interaction. In line with this conception, we present a taxonomy of human interaction patterns; then propose five different feature sets, comprising force-, velocity- and power-related information, for the classification of these patterns. Our evaluation shows that using a multi-class support vector machine (SVM) classifier, we can accomplish a correct classification rate of 86 percent for the identification of interaction patterns, an accuracy obtained by fusing a selected set of most informative features by Minimum Redundancy Maximum Relevance (mRMR) feature selection method.

Index Terms—Behavior recognition; classifier design and evaluation; feature evaluation and selection; haptic collaboration; haptic interfaces; haptics-enabled virtual environments; interaction patterns; machine learning; pattern recognition; physical human-X interaction; realistic haptic human-robot interaction; support vector machine classification

1 Introduction

With the emergence of the idea of autonomy in the robotics domain, a significant amount of research has shifted towards discovering how to make robots act in a more human-like manner in terms of their social, cognitive, and motor abilities. Significant attention is now directed towards building interactive and proactive robotic systems, which are capable of cooperating with humans in everyday situations instead of assisting with specific and possibly industrial tasks. In order to build cooperative robotic systems that allow natural and intuitive interaction, an understanding of human behavior and intentions, as well as a capability for communication and coordination is required. In this paper, we follow a human-centric experimental approach to discover human behavior characteristics in everyday physical tasks. We believe that the information extracted from the operation of two humans is invaluable for developing a robotic *partner* that can effectively cooperate with humans.

Humans cooperate through numerous physical activities during their daily routines. These activities cover a wide range of tasks, such as jointly moving objects, assembling machine parts, hand shaking, and dancing. In its broader sense, cooperation addresses interaction characteristics that provide mutual benefit to the partners. Thus we expect partners to work in harmony or at least without inhibiting the natural course of a given task. However, from time to time, the continuous and complex nature of physical tasks may necessitate partners to adopt some non-cooperative behaviors (i.e. conflicts). Imagine a couple, which has trouble in synchronizing their movements while dancing waltz. The conflict they face can be solved as soon as they manage to move along with the music simultaneously. Such conflicts -unintentional as they are- may be due to differences in partners' intentions or discrepancies in

Fig. 1. Daily collaboration scenario: two humans jointly carry a table.

- *C. E. Madan, T. M. Sezgin and C. Basdogan are with the College of Engineering, Koc University, Sariyer, 34450, Istanbul, Turkey. E-mail: cmadan, mtsezgin, cbasdogan@ku.edu.tr*
- *A. Kucukyilmaz is with the Electrical and Electronics Department, Imperial College London, Exhibition Road, SW7 2BT, London, UK. E-mail: a.kucukyilmaz@imperial.ac.uk*

reaction times to each other's actions. Determining how and when interaction behaviors change is a key issue in understanding human collaboration.

A robot, which can comprehend how humans interact, would be able to either mimic the behaviors of one of the partners, or complement the interaction of humans as an assistant. As a motivating example, think of a robotic system that aids two people with the installation of a rooftop car rack. The humans stand on both sides of the car and try to balance the rack in the correct pose while the robotic system helps them with carrying the heavy load. In this case, humans do not act as dyads just because they need help from one another, but because dyadic interaction becomes the medium of communication. In this example, assume that the robot is not fitted with tools to determine where the rack should be installed, but is only capable of lifting the rack up or down as well as monitoring the interaction between humans. In this scenario, the task needs to be led by the humans. However the robotic system can effectively help in completing the task by speeding up the operation in the right direction when it recognizes harmony between the partners and stabilizing the rack when it infers a conflict between them. In other words, the robotic system recognizes the interaction patterns of the humans partners, and assists them as needed.

This study is an effort to investigate interaction patterns in human-object-human scenarios, where two humans cooperate to move an object (see Fig. 1). We focus on dyadic joint object manipulation tasks to identify human interaction patterns when partners collaborate in the existence of conflicts.[1] In this sense, this study is a first step towards exploring how the partners' intentions change over the interaction in a physical task. In order to observe the interaction patterns of the partners, we designed four different dyadic object manipulation scenarios in a haptics-enabled virtual environment. Two of these scenarios were designed to promote collaboration between the partners without imposing any conflict on them, while the other two artificially invoke conflicts between the partners. Real human-human interaction data is collected through a controlled user study with 20 dyads. Through offline examination of this data, we observed that partners exhibit specific interaction patterns during joint operation. Specifically, we first defined three possible interaction types (harmonious, conflicting, and neutral) and then identified six interaction patterns based on the intentions of the dyad on the motion. A human annotator observed the video recordings of the trials, which were captured during the experiment, and manually annotated the data. Through this process, we identified the meaningful parts of the collected data and labeled them with the aforementioned interaction patterns to form a labeled set of data for supervised learning.

We conducted a set of statistical analyses on the data in order to find descriptive variables that are used to recognize the interaction patterns. These descriptive variables are: 1) forces applied by individual agents on the manipulated object, 2) net force applied by the partners on the manipulated object, 3) interactive force among the partners, 4) velocity of the manipulated object, and 5) power transferred to the manipulated object by the partners. We formed five different feature sets, four of which are composed of haptic information, by extracting features from these descriptive variables. For the recognition of interaction patterns, we used multi-class support vector machine (SVM) classifiers with these 5 feature sets. The classification results indicate that each individual feature set was successful in recognizing at least 4 of the 6 interaction patterns.

Even though the individual feature sets fail to recognize all interaction patterns, when the features are fused to obtain an optimal feature set by the Minimum Redundancy Maximum Relevance (mRMR) feature selection method, we can accomplish a correct classification rate of 86 percent for the identification of interaction patterns.

This paper is organized as follows: Section 2 presents the related work. The experimental setup used for data collection is described under Section 3. The interaction patterns observed during the experiment, and the proposed taxonomy are discussed in Section 4. The machine learning method that is used for the classication of interaction patterns is explained in Section 5. The results and the discussion are presented in Section 6, finally followed by conclusions in Section 7.

2 Background

Developing robots that can collaborate with human partners during physical interaction requires the robots to display proactive behavior. So far, the widespread approach to realize proactive behavior has been to improve the control schemes of the robots based on estimations of human intentions. In an early study, Rahman et al. programmed the robot to replay task-specific trajectories recorded in human-human experiments to generate human-like velocity trajectories in human-robot cooperation [22]. Later, Tsumugiwa et al. estimated human arm stiffness through the observation of measured position and forces, and adapted the admittance parameters accordingly [27]. Similarly, Duchaine and Gosselin implemented variable admittance based on the velocity and force derivative information obtained from the human [6]. Corteville et al. developed a human-inspired robotic assistant, which assumed that the humans follow a minimum jerk trajectory [9] during motion, and estimated the intended motions of the human partner based on his/her position and velocity profile [5]. The robot then adjusted its velocity profile to fit along with the intended velocity.

Alternatively, some other investigators have focused on role allocation and sharing in human-robot interaction. Evrard and Kheddar defined two distinct extreme behaviors (leader and follower) for partners and switched between the behaviors via two independently-varying functions [8]. Later, Kheddar illustrated the use of this mechanism during collaboration with a humanoid robot [13]. Similarly, Bussy et al. proposed a control law for physical interaction with a

1. Note that even though we focus on dyadic interaction in this paper, the ideas we present can be easily extended to multi-agent scenarios.

humanoid robot in an object transportation task [3]. Their control law enabled the robot to proactively switch between standalone (i.e. performing the task alone) and collaborative (i.e. leader or follower) roles depending on the intentions of its human partner. Oguz et al. [20] and Kucukyilmaz et al. [15], [16] proposed a method to infer the intentions of the human during a joint object manipulation task. They implemented a dynamic role exchange model, where the robot inferred human's intentions based on the forces applied by him/her, and chose between leader or follower roles. Later, Mörtl et al. presented a similar dynamic role exchange mechanism for a joint object manipulation task, in which a man-sized mobile robot sensed the human partner's intentions through the evaluation of an agreement criterion based on the human's force input, and helped accordingly [19]. These studies enhanced human-robot interaction via generating more natural trajectories. However, the rule based nature of the control laws utilized in these studies makes them difficult to generalize for different tasks. Furthermore, even though the robots are capable of adapting to their human partners, they lack the ability to comprehend how human behaviors change during interaction, and what these changes signify.

A widely accepted perspective advocates the investigation of human-human interaction to learn from the behavioral mechanisms utilized by humans. Based on the insight gained from human-human interaction data, Reed and Peshkin illustrated that two opposing intentions, to accelerate or to decelerate, exist in a dyadic target acquisition task [23]. Similarly, Stefanov et al. specified conductor and executor roles, which bear information about how two humans cooperate in a joint manipulation task [25]. They presented a model based on velocities and interaction forces to define the roles. Groten et al. focused on the consistency of dominance behavior during a tracking task where two humans collaborated [10]. They demonstrated that the participants' interaction can be represented by a personal dominance distribution. Later, they investigated how partners communicate through intentions, and suggested that in order to achieve a joint goal, partners need to integrate their individual action plans in both collaborative and conflicting situations [11]. Even though these studies adopt a similar approach to that of ours, in a sense that they examine human-human interaction data, they are inherently different. All these studies focus on presenting the existence of different patterns in human behaviors; however, none of them attempt a systematic classification of such patterns using machine learning techniques. Additionally, they mainly define individual labels for human intentions, but do not focus on how partners work with each other over time.

In order to address this shortcoming, some researchers have used statistical learning models to infer about the intentions of the human partner. Evrard et al. implemented a learning-by-demonstration technique [2] to differentiate between leader and follower roles [7]. Their system was able to capture the role switching moments using Gaussian Mixture Models. Takeda et al. [26] and Wang et al. [28] proposed HMM based algorithms to estimate human intentions in physical dyadic tasks, where a robot collaborated proactively with its human partner. Schrempf et al. presented a new approach that allows a robot to plan its actions even if the human intention estimation was uncertain [24]. In their system, the robot computed a confidence for possible actions and executed the task by selecting actions proactively. Carlson and Demiris defined certain actions that can be performed while driving a powered wheelchair, then dynamically predicted the most probable actions that shall be taken in the near future [4]. Even though these studies presented task-independent solutions to intention recognition, they fell short in interpreting the meaning of the intentions and the interaction patterns.

Characterization of interaction patterns is an emerging topic in human-human and human-robot interaction domains. As the name implies, interaction patterns describe the interaction between agents, not the behavior of an individual. In this sense, it provides a different perspective to the same problem. There are a few studies in literature, which identify a taxonomy of interaction patterns and perform task-dependent classification. Recently, Jarrasse et al. have introduced a general taxonomy of interaction patterns in physical tasks [12]. They formulated the human-robot interaction patterns as controllers. The proposed framework provided a description of interaction patterns of a dyad executing a joint task, along with an interpretation of the patterns. Even though the utility of this taxonomy was demonstrated by simulated interactions of two humans, it lacked the identification of patterns in real data. Melendez-Calderon et al. defined five human interaction patterns in a tracking task where two humans worked together [18]. The patterns were defined as templates, which described the action of each partner, such as one agent accelerating the movement while the other is braking. They proposed a rule-based classification system using the interaction torques and EMG recordings of partners' activities to identify these patterns. However, their technique was highly task-dependent. Besides, it required manual construction of templates and a lot of fine tuning when the task dynamics changes. Furthermore, the system was not robust against the addition of new interaction strategies.

Even though the aforementioned studies provide valuable information about human interaction patterns, to our knowledge, no effort has yet been put into building a systematic way of defining and recognizing these patterns. In this sense, our work is a first to both present a taxonomy and propose a recognition framework for real human-human interaction data. Additionally, the classification method proposed in this paper aims at discovering the descriptive features of interaction, hence, given training data, our technique can be applied to a diversity of tasks.

3 EXPERIMENT

We conducted an experimental study to generate data that can be used to identify human-human haptic interaction patterns and learn models for capturing salient charac-

(a) Screens of Agent 1 and 2

(b) Agent 1 and 2

Fig. 2. Two humans interact through haptic devices in order to jointly move an object, resembling a table with caster wheels, in a virtual environment.

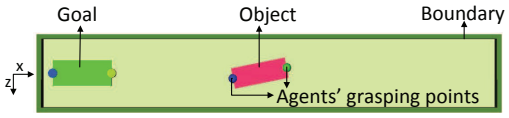


Fig. 3. The straight scene

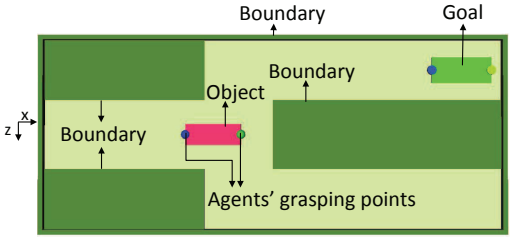


Fig. 4. The bifurcated scene

teristics of dyadic interactions.[2] This section presents the experimental design and the scenarios used in this study, as well as the physics-based engine underlying the virtual environment and the experimental procedures.

3.1 Experimental Environment

In order to identify human interaction patterns, we have developed an application where two human subjects interact in a virtual environment through the haptic channel. During the experiment, the subjects were situated in different rooms, so that they only interact through haptic devices.

The application requires the subjects to coordinate their actions in order to move the rectangular object together in a 2D maze-like scene (see Figs. 2). Due to the selection of friction coefficients, the object rotates easily within the environment, resembling the motion of a table moving on caster wheels. The goal of the task is to move the object toward a target parking configuration and stay there for a predetermined period of 5 seconds.

During the experiment, the subjects are presented with two different scenes to observe interaction patterns in both translational and rotational motion. The first scene, which will be called the *straight scene* from now on, depicts a horizontal path, whereas the second scene, called the *bifurcated scene*, presents a fork-shaped path for the users to follow. Obviously, the straight scene involves translation along a straight line, while the bifurcated scene entails both translation and rotation. Screenshots of the application for each scene can be seen in Figs. 3 and 4.

As seen in the screen visuals, the jointly manipulated object is depicted as a pink rectangular block. The grasping points of agents are represented as blue and green spheres attached to its short edges. The target is visually represented with a green rectangle that resembles the object and clearly conveys the desired final orientation. Once the object reaches the desired configuration, the target turns blue and a counter appears in the middle of the screen to alert the user. If the agents succeed in staying on the target until the end of the countdown, a new target appears somewhere else in the scene. In both scenes, boundaries constrain the movement of the object. Hitting the boundaries during the task is considered an error. In order to signal such errors to the users, the boundaries turn red on collision.

3.2 Physics-Based Engine

This section details the physics-based engine underlying the virtual environment. Note that bold-face symbols are used to denote vectors throughout the section.

The manipulated object is modeled as a rigid body that moves in 2D in a way similar to the movement of a table moving on caster wheels. The physics based engine conveys the dynamic nature of the task to the agents both visually and through haptics. The agents interact with the environment via haptic devices. The end-effector positions of haptic styli along x- and z-axes map to the positions of the individual *haptic interface points* (HIPs). A spring and damper model is used between each agent's HIP and the grasping point on the object, as shown in Fig. 5. The model is used to calculate the individual forces applied by the agents on the object:

$$\boldsymbol{F_{HIP_1}} = K_p(\boldsymbol{x_{HIP_1}} - \boldsymbol{x_{g_1}}) + K_d(\dot{\boldsymbol{x}}_{\boldsymbol{HIP_1}} - \dot{\boldsymbol{x}}_{\boldsymbol{g_1}}) \quad (1)$$

$$\boldsymbol{F_{HIP_2}} = K_p(\boldsymbol{x_{HIP_2}} - \boldsymbol{x_{g_2}}) + K_d(\dot{\boldsymbol{x}}_{\boldsymbol{HIP_2}} - \dot{\boldsymbol{x}}_{\boldsymbol{g_2}}) \quad (2)$$

where K_p and K_d are spring and damper coefficients, $\boldsymbol{x_{HIP_1}}$, $\boldsymbol{x_{HIP_2}}$, $\dot{\boldsymbol{x}}_{\boldsymbol{HIP_1}}$, $\dot{\boldsymbol{x}}_{\boldsymbol{HIP_2}}$ are the positions and velocities of HIPs, and $\boldsymbol{x_{g_1}}$, $\boldsymbol{x_{g_2}}$, $\dot{\boldsymbol{x}}_{\boldsymbol{g_1}}$, $\dot{\boldsymbol{x}}_{\boldsymbol{g_2}}$ are the positions and velocities of the grasping points of the agents. Reciprocally, the agents are fed back with forces $-\boldsymbol{F_{HIP_1}}$ and $-\boldsymbol{F_{HIP_2}}$ through the haptic devices, so that they can feel the dynamics of the object.[3]

Note that this design utilizes equal gains for the spring-damper model of each agent.[4] This implies that the agents

2. The raw data generated through this experiment and the labeled dataset of annotated interaction segments are publicly available through http://rml.ku.edu.tr/HHIBehaviorDataset.

3. Due to mechanical constraints of the haptic devices, the forces fed back to the humans are thresholded at 4.0 N.

4. The values of the spring and damper coefficients were respectively set to $K_p = 0.25$ N/mm and $K_d = 0.001$ Ns/mm.

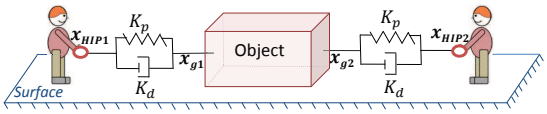


Fig. 5. HIPs are connected to the object with spring/damper systems. K_p and K_d are spring and damper coefficients

are of assumed to be of equal strength. However, it is worth noting that this design choice does not prevent the partners from applying different forces on the object during the interaction by moving the end-effectors of their haptic devices. In fact, discrepancies in interaction dynamics are naturally manifested through the way the agents apply forces on the object.

In addition to the applied forces, in case the object collides with the boundaries, an impact force, $\boldsymbol{F_I}$ is applied on the object to prevent penetration of the object into the boundaries. Furthermore, since the object acts as a rigid body, moments are generated due to the forces applied on the object by the agents ($M_{HIP_u}, u = 1, 2$) and due to the impact force (M_I). For simplicity, 2D dynamics is assumed and the direction of momentum is considered to be always orthogonal to the movement plane.

The object is also affected by frictional forces due to its contact with the surface. Translational and rotational friction ($\boldsymbol{F_f}$ and M_f) are calculated using the Coulomb friction model.[5] Thus, the net force and moment acting on the object becomes:

$$\boldsymbol{F_{net}} = \boldsymbol{F_{HIP_1}} + \boldsymbol{F_{HIP_2}} + \boldsymbol{F_I} + \boldsymbol{F_f} \quad (3)$$

$$M_{net} = M_{HIP_1} + M_{HIP_2} + M_I + M_f \quad . \quad (4)$$

The state of the object at each time step $(\boldsymbol{x_{obj}}, \dot{\boldsymbol{x}}_{\boldsymbol{obj}}, \Theta_{obj}, \dot{\Theta}_{obj})$ is calculated from M_{net} and $\boldsymbol{F_{net}}$ using Euler integration.

3.3 Scenarios

In order to elicit different interaction patterns, we presented the subjects with different manipulation scenarios, in which conflicts between partners are artificially invoked by providing each agent with different visual information about the location of the target configuration. Apart from the target locations, both subjects observe the motion of the object and view the same path. The subjects are not aware of the whereabouts of their partner's target, but they are informed that it can be different from that of their own, or either they or the other agent might not be given a target at all. Note that the scenarios do not force partners to act in a well-defined and straightforward manner. Instead, collaborating agents can display different behaviors during interaction, which are shaped with respect to the characters and emotional states of the individuals (See [29] for a thorough discussion on this). The scenarios are designed only to improve the diversity of the resulting behaviors.

The following manipulation scenarios are considered in the experimental study:

Scenario 1: Harmony

In this scenario, both subjects are given the same target. Hence, we expect no conflict in terms of final goals. Fig. 6(a) represents the screen visual shown to each subject for both straight and bifurcated scenes.

Scenario 2: Full Conflict

The subjects are presented with conflicting goals in this scenario. The target configurations are arranged so that only one of them can be achieved at the end of the task. As a result, one of the subjects needs to yield to the authority of the other in order to accomplish the task. Fig. 6(b) shows the screen visual shown to each subject for both scenes.

Scenario 3: Partial Conflict

Similar to the previous scenario, conflicting targets are given to subjects. The achievement of both tasks is not possible, yet the conflict manifests itself later during the trial and the amount of conflict is expected to be less than that of Scenario 2. Fig. 6(c) represents the screen visual shown to the subjects for both scenes.

Scenario 4: Single Blind

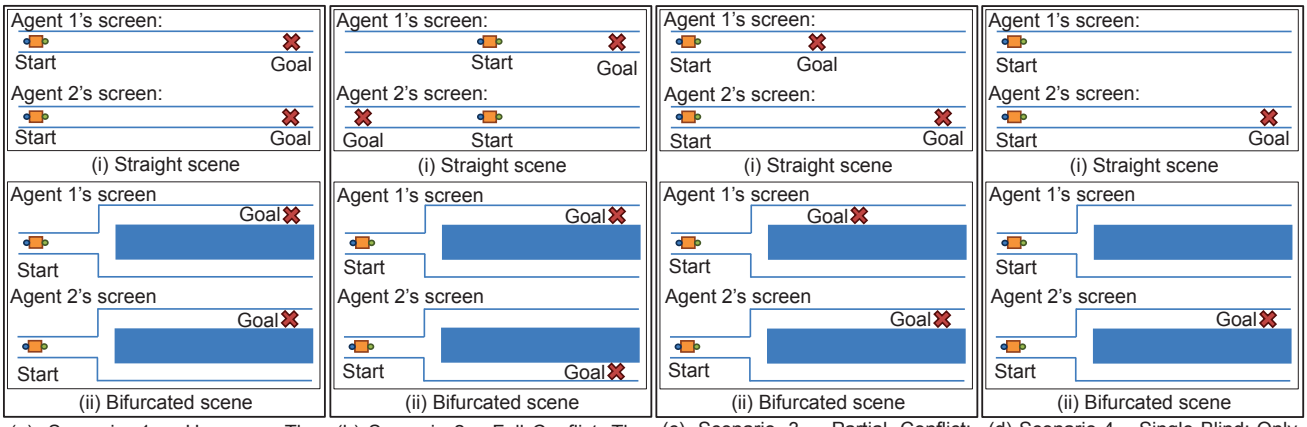
In this scenario, only one subject is assigned a goal. The other subject (i.e. the blinded subject) is informed that (s)he needs to observe and follow the actions of his/her partner. It is possible to accomplish the task, but the blinded subject is expected to get confused. Fig. 6(d) represents the screen visual shown to the subjects for both scenes. Note that in this figure, the blinded subject is agent 1, however a dual scenario, where agent 2 acts as the blinded subject, is also considered in the experiments.

3.4 Procedure and Participants

40 subjects (6 female and 34 male), aged between 21 and 29, participated in our study. The subjects were randomly divided into two groups to form dyads that should work as partners during the experiment. The partners were separated in two different rooms, so that they could not see or hear each other. They interacted with the object and each other through Geomagic®(formerly Sensable®) Phantom®Premium™haptic devices using a stylus attachment. The haptic devices were connected to separate PCs and communicated through a UDP connection over the local area network.

At the beginning of the experiments, each participant was presented with the same goals (i.e. Scenario 1) for two practice trials in order to familiarize him/her with the system. During the experiments, each manipulation scenario was presented twice, hence, there were a total of

5. The values of the static and kinetic friction coefficients were respectively set to $\mu_{t,s} = 0.19$ and $\mu_{t,k} = 0.15$ for the translational case and to $\mu_{r,s} = 0.20$ and $\mu_{r,k} = 0.19$ for the rotational case.



(a) Scenario 1 - Harmony: The agents are provided with the same goals, which lie at the farthest end of the path.

(b) Scenario 2 – Full Conflict: The agents have conflicting goals, which lie (i) at opposing ends of the corridor for the straight path and (ii) at the end of different branches of the bifurcated path.

(c) Scenario 3 - Partial Conflict: Both agents' goals are on the same path, however one of the agent's goal is closer.

(d) Scenario 4 – Single Blind: Only one agent is provided with a goal at the farthest end of the path, whereas the other agent does not see any goal on screen.

Fig. 6. Four scenarios in straight and bifurcated scenes.

10 trials[6] to be analyzed. In order to balance the learning effects, the order of the scenarios were permuted using a Latin square design. The subjects were not given detailed descriptions of the scenarios or the interaction patterns, but they were informed that their partners may have conflicting goals or no goal at all.

4 A Taxonomy of Human Interaction Patterns

Based on our interpretations of user interactions after the experiments, we identified a set of interaction patterns that were observed frequently in our dyadic object manipulation task. These constitute our taxonomy of human interaction patterns as illustrated in Fig. 7. In the proposed taxonomy, the first layer presents a very general categorization of any physical interaction involving multiple agents. In this layer, an interaction-based perspective is adopted to classify the task as being either harmonious, conflicting, or neutral. The second layer is concerned with the "intentions" of the agents. In this sense, it is not related to the resulting motion of the object itself, but is rather responsible for defining whether the agents' motion plans agree or not. Finally, the last layer describes interaction patterns that are commonly encountered in our task. These patterns can be explained within the scope of the proposed taxonomy as follows:

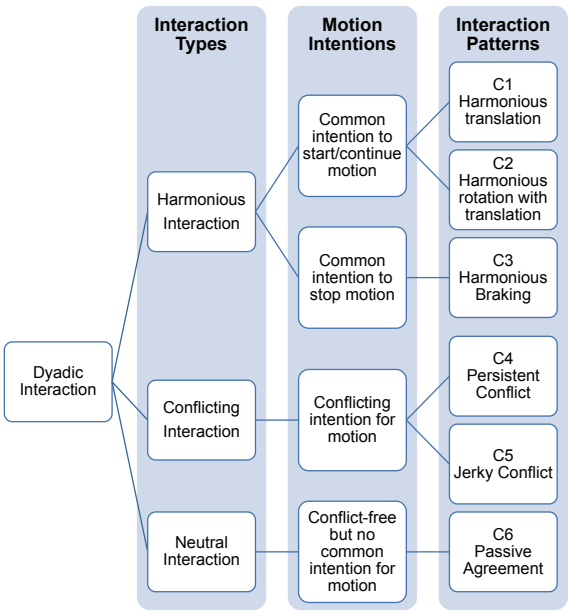


Fig. 7. Taxonomy of interaction patterns in dyadic object manipulation

1. Harmonious interaction:

The partners move the object while agreeing on the direction of the movement. In other words, the intention of both agents are the same; thus, no conflict exists between the agents. We examine this interaction type in two subclasses:

a) **Common intention to start/continue motion:** The manipulated object accelerates or moves with a constant velocity.
   i) **Harmonious translation (C1):** The partners agree on translating the object. In other words, both agents apply forces in the same direction to translate the object.
   ii) **Harmonious rotation with translation (C2):** The partners voluntarily rotate the object by agreeing on moving it along an arc or about its center.

b) **Harmonious braking (C3):** One or both partners voluntarily decelerate the object with the purpose of stopping the motion. In practice, at least one agent starts applying force in the direction opposite to the movement until the object is stationary.

6. Note that Scenario 4 was presented in a twofold fashion so that each agent gets to act as the blinded user in the experiment.

2. Conflicting interaction:

The interaction is dominated by some form of conflict between the agents. In other words, the partners have no common intention for motion. In this type of interaction, we expect that the partners can neither move the object smoothly nor achieve their goal. Two patterns can be defined for this interaction type:

i) **Persistent conflict (C4):** The partners insist on moving the object in opposite directions and hence the object does not move much.

ii) **Jerky conflict (C5):** The users disagree on the movement of the object, but not in a persistent fashion. This typically causes the object to rotate involuntarily or follow undesired trajectories, possibly ending with collisions with the environment. In more general terms, this pattern can be thought of any apparent conflict between agents that is not persistent.

3. Neutral interaction:

This interaction type implies no conflict between the partners. However, the agents share no common intention for the motion, either. Interaction is mainly governed by an agent being passive, which defines a single interaction pattern:

i) **Passive agreement (C6):** At least one of the partners remains passive by not contributing much to the task.

5 Recognition of Human Interaction Patterns

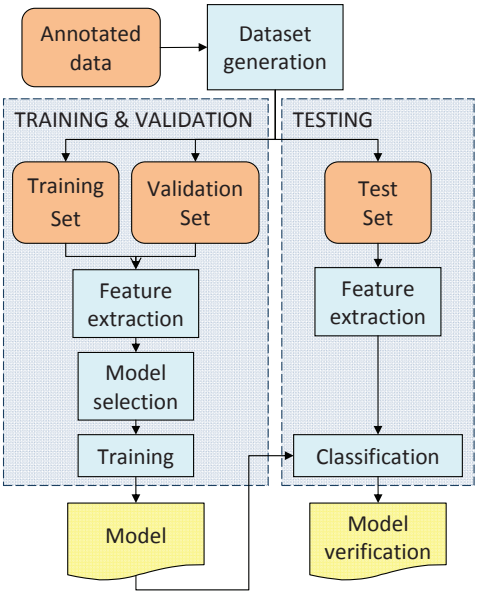


Fig. 8. Stages of classifier learning.

Our statistical pattern classification system possesses the structure given in Fig. 8. First, raw data is annotated by an annotator, who has a good understanding of the taxonomy and the interation behaviors, to obtain a set of meaningful labeled interaction segments. Then, in order to avoid overfitting, the data is split into 3 distinct parts, namely training, validation, and test sets by randomly selecting instances from the whole data. The training and validation sets are used to estimate parameters of the classifier, while the test set is used to assess the performance of the fully trained classifier. In particular, we initially divide the data into two by hold-out cross validation. This process separates our test set by setting aside 30 percent of the data. This guarantees that the test set is only used to assess the learned model. The remaining 70 percent is then divided into training and validation sets via 5-fold cross validation in order to select the best parameters and train the model using a separate set. After the creation of the datasets, features are extracted for each of them, and model training is performed using only the training and validation sets. Once the SVM is trained with the optimal parameters, it is used for the classification of patterns using the test set. The steps of our learning procedure is as follows:

1) Annotate raw data
2) Divide the data into training, validation, and test sets
3) Extract features from training, validation, and test sets
4) Select model parameters
5) Train the model using the training set
6) Evaluate the model using the validation set
7) Repeat steps 4 - 6 with different model parameters
8) Select the best parameters and train the model using the training and validation sets
9) Assess the final model using the test set

The details of these stages will be explained in the rest of this section.

5.1 Annotation of Interaction Pattens

After the experiment, we generated videos of the trials by simulating the recorded data in Matlab® environment. Regarding the videos, the data is manually annotated with the interaction behaviors using the ELAN annotation tool [1]. The annotation is performed by a human, who has a thorough understanding of the proposed taxonomy and the interaction behavior classes.[7]

At the end of the annotation process, variable-length labeled interaction segments were obtained. Segments shorter than 4 seconds were discarded to eliminate the noise due to instantaneous behavior changes during interaction. As a result, we obtained a populated dataset of 1944 instances. The percentage of instances per interaction pattern class is shown in Fig. 9. The number of instances are particularly small in harmonious rotation with translation (C2), harmonious braking (C3), and persistent conflict (C4) classes. Upon examining the dataset, we observed that these classes of behaviors are indeed encountered infrequently during

7. The reliability of the annotation process is crucial for the accuracy of any supervised learning task and it is possible that the data annotated by a single annotator may suffer from human-error and subjectivity. In order to validate the primary annotator's reliability in labeling the behaviors, we instructed another annotator to independently perform the annotations. Then, an inter-rater agreement analysis was conducted to observe whether the resulting annotations are consistent across different annotators. As a result, a Krippendorff's alpha value of 0.91, which is high enough to indicate the consistency of the behavior definitions, is computed [14].

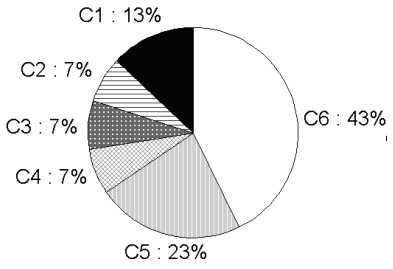


Fig. 9. Percentage of instances per interaction pattern class in the dataset. C1: Harmonious translation, C2: Harmonious rotation with translation, C3: Harmonious braking, C4: Persistent conflict, C5: Jerky conflict, C6: Passive agreement

the experiment. A possible explanation for this can be as follows: C2 is only required in the bifurcated scene at or near the bifurcation region. Since any rotation caused by conflicts is considered to be either in C5 or C6 classes, C2 is less populated than the others. Similarly, C3 is encountered typically at the end of the trials only once, when the users attempt to park the object. Finally, the frequency of C4 is low, since it is mostly encountered in full or partial conflict scenarios, and resolved easily.

5.2 Identification of Meaningful Features

The success of any pattern recognition system relies on the presence of informative features. At the end of the annotation process, we obtain a bulk set of labeled data, consisting of the agents' forces as well as variables related to the movement of the object, such as its position, orientation, linear and angular velocity, and acceleration. Prior to classification, in order to have an initial idea about the descriptive power of the these variables, we ran ANOVAs to seek differences between class means for each variable.

Fig. 10 shows the means and the standard errors of means for each pattern class for each of these variables. Statistically significant effects ($p < 0.001$) are detected for all the variables. However, it is important to note that statistically significant differences between classes do not necessarily imply high recognition accuracies during classifications. The predictive classification accuracies for each feature set are further discussed in Section 6.

5.2.1 Mean Magnitude of the Individual Forces Applied by the Agents

Individual forces exerted by the subjects ($\boldsymbol{F_{HIP_1}}$ and $\boldsymbol{F_{HIP_2}}$) are averaged over the duration of the interaction:

$$MF_{HIPs} = \frac{1}{2T}\sum_{u=1}^{2}\int_T \|\boldsymbol{F_{HIP_u}}\|\mathrm{d}t \quad , \tag{5}$$

where T is the length of the interaction sequence.

5.2.2 Mean Magnitude of the Net Force Applied by the Agents

The net force is the vector sum of the agent forces applied on the manipulated object. The mean magnitude of the net force exerted by the agents is calculated by:

$$MF_{net} = \frac{1}{T}\int_T \|\boldsymbol{F_{HIP_1}} + \boldsymbol{F_{HIP_2}}\|\mathrm{d}t \quad . \tag{6}$$

5.2.3 Mean Magnitude of the Interactive Force Acting on the Object

The interactive force, f_i acting on the object reflects the internal force that acts on the object. Interactive force is defined in the redundant task space [17] and occurs if the agents apply "compressive or tensile forces that do not contribute to the motion of the object" [10]. Interactive force is defined as:

$$f_i = \begin{cases} F_{HIP_{1x}} & \text{sign}(F_{HIP_{1x}}) \neq \text{sign}(F_{HIP_{2x}}) \\ & \wedge |F_{HIP_{1x}}| \leq |F_{HIP_{2x}}| \\ -F_{HIP_{2x}} & \text{sign}(F_{HIP_{1x}}) \neq \text{sign}(F_{HIP_{2x}}) \\ & \wedge |F_{HIP_{1x}}| > |F_{HIP_{2x}}| \\ 0 & \text{sign}(F_{HIP_{1x}}) = \text{sign}(F_{HIP_{2x}}) \end{cases} , \tag{7}$$

where $F_{HIP_{1x}}$ and $F_{HIP_{2x}}$ stand for the x components of the agent's applied forces in the object frame. The mean magnitude of the interactive force acting on the object (MF_i) is calculated as:

$$MF_i = \frac{1}{T}\int_T |f_i|\,\mathrm{d}t \quad . \tag{8}$$

5.2.4 Mean Magnitude of the Linear Velocity of the Object

The mean magnitude of the linear velocity of the object is calculated as follows:

$$M\dot{x} = \frac{1}{T}\int_T \|\dot{\boldsymbol{x}}_{\boldsymbol{obj}}\|\mathrm{d}t \quad . \tag{9}$$

5.2.5 Mean Magnitude of the Angular Velocity of the Object about the y-axis

The mean magnitude of the angular velocity of the object about the y-axis ($\dot{\theta}_{obj}$) is calculated as follows:

$$M\dot{\theta} = \frac{1}{T}\int_T \left|\dot{\theta}_{obj}\right|\mathrm{d}t \quad . \tag{10}$$

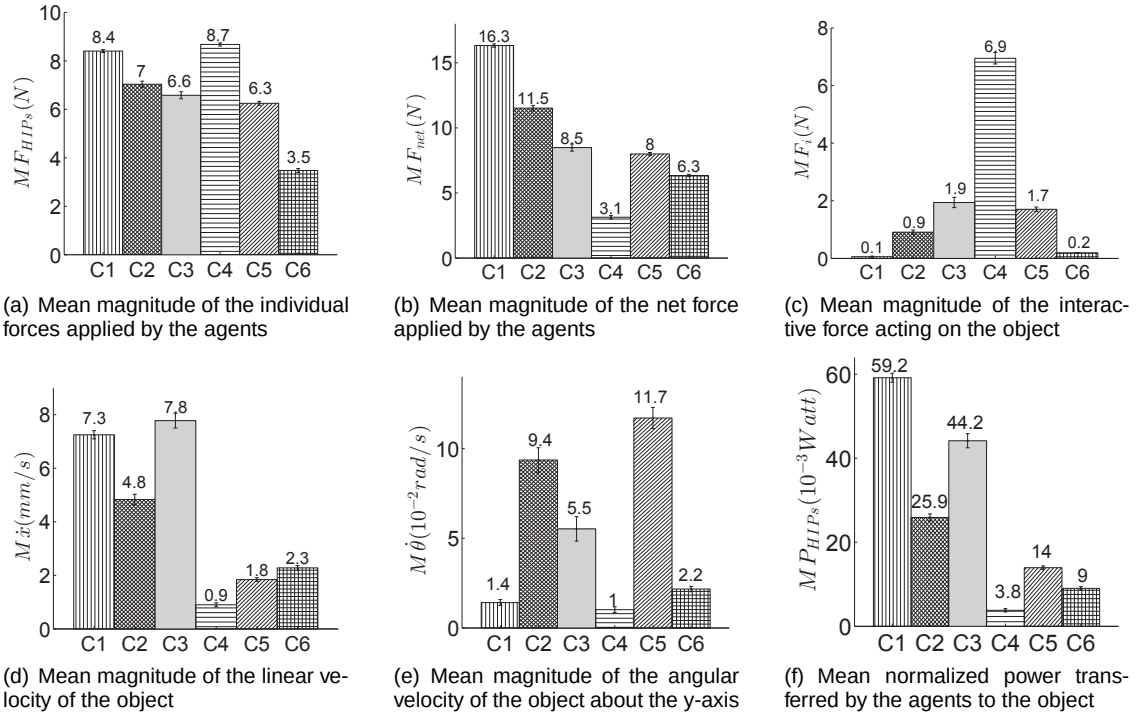
5.2.6 Mean Normalized Power Transferred by the Agents to the Object

The power transferred by agents to the object is calculated as follows:

$$P_{HIP_u} = \int_T \left(|\boldsymbol{F_{HIP_u}} \cdot \dot{\boldsymbol{x}}_{\boldsymbol{obj}}| + \left|M_{HIP_u}\dot{\theta}_{obj}\right|\right)\mathrm{d}t, \tag{11}$$

where $u = 1, 2$. Using this, the mean normalized power transferred by the agents to the object (MP_{HIPs}) is calculated as:

$$MP_{HIPs} = \frac{1}{2T}\sum_{u=1}^{2} P_{HIP_u} \quad . \tag{12}$$



(a) Mean magnitude of the individual forces applied by the agents

(b) Mean magnitude of the net force applied by the agents

(c) Mean magnitude of the interactive force acting on the object

(d) Mean magnitude of the linear velocity of the object

(e) Mean magnitude of the angular velocity of the object about the y-axis

(f) Mean normalized power transferred by the agents to the object

Fig. 10. Mean values of variables for each pattern class. The error bars indicate standard errors of the means. C1: Harmonious translation, C2: Harmonious rotation with translation, C3: Harmonious braking, C4: Persistent conflict, C5: Jerky conflict, C6: Passive agreement

5.3 Dataset Generation and Feature Extraction

The annotation process results in variable length interaction segments. However, in order to be used in classification, we need to represent the data using a fixed number of features for each annotated interaction segment. In order to come up with the most informative features, we followed a systematic subdivision approach, which divides the whole interaction segment into support regions and then computes the mean, standard deviation, median, and interquartile range values for each region. This is motivated by the idea that some behaviors are not consistent throughout the interaction segment and it is not possible to capture those behaviors using descriptive statistics from the whole interaction segment. For instance, during braking, we expect the magnitude of the velocity to drop over time. In order to capture such temporal properties, we divide the whole interaction segment into 12 support regions as shown in Figure 11. The support regions cover the following ranges:

$$\begin{aligned}
R1 &: [0 \quad T] \\
R2 &: [0 \quad \Delta t] \\
R3 &: [T-\Delta t \quad T] \\
R4 &: [T/2-\Delta t \quad T/2+\Delta t] \\
R5, R6 &: [iT/3-\Delta t \quad iT/3+\Delta t], i=1,2 \\
R7-8 &: [iT/4-\Delta t \quad iT/4+\Delta t], i=1,3 \\
R9-12 &: [iT/5-\Delta t \quad iT/5+\Delta t], i=1,2,3,4 \quad ,
\end{aligned}$$

where $\Delta t = 1$ s. The number of support regions is selected empirically to span the whole interaction segment as much as possible without inflating the total feature count.

Computing 4 features (mean, standard deviation, median, and interquartile range) for each variable separately over the aforementioned 12 support regions, 5 datasets are constructed. Specifically, Set 1 is related to the interaction forces, which are measured in 2-dimensions for each agent; hence it contains 192 features (12 support regions x 4 features x 2 dimensions x 2 agents). Set 2 is related to the net forces and stores 96 features (12 support regions x 4 features x 2 dimensions); Set 3 is related to the interactive forces and stores 48 features (12 support regions x 4 features); Set 4 is related to the linear and angular velocities of the object and stores 144 features (12 support regions x 4 features x 3 dimensions); and finally Set 5 is related to the power consumed by agents and stores 96 features (12 support regions x 4 features x 2 agents). Table 1 presents the feature sets used in this study. Each row of this table defines a separate feature set, which will be assessed for its discriminative power.

TABLE 1
Feature sets

Set	Set Name	Features	Count
Set 1	Agent force-related	$\boldsymbol{F_{HIP_1}}$, $\boldsymbol{F_{HIP_2}}$	192
Set 2	Net force-related	$\boldsymbol{F_{net}}$	96
Set 3	Interactive force-related	f_i	48
Set 4	Velocity-Related	$\dot{\boldsymbol{x}}_{\boldsymbol{obj}}$, $\dot{\theta}_{obj}$	144
Set 5	Power-Related	P_{HIP_1}, P_{HIP_2}	96
Total			576

5.4 Classifier Design

We utilize a multi-class Support Vector Machine (SVM) classifier with a Gaussian radial basis function kernel to

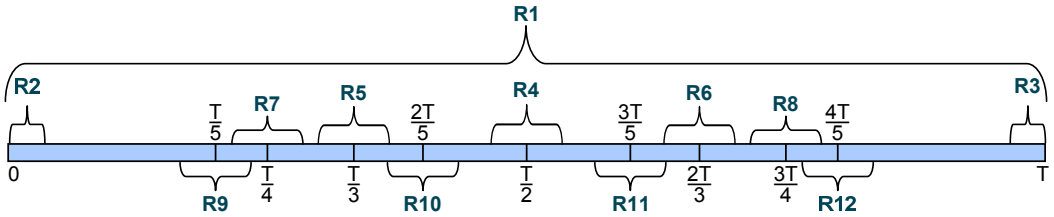


Fig. 11. Twelve support regions are constructed through systematic subdivision of the whole interaction. Equal number of features are computed over each region.

recognize interaction patterns. In order to deal with the multi-class learning problem, we adopt the one-against-one strategy, which builds one SVM for each pair of classes.In order to obtain the optimal hyper-parameters, cost and γ of the model, we perform model selection by 5-fold cross-validation using grid search.

5.5 Evaluation

For the evaluation of the classifier performance, we utilize the following metrics:

1) **Normalized Confusion Matrix** is a table which displays the correct and incorrect classification rates of each class. The values in the columns and rows respectively represent the number of instances in the predicted and the actual classes normalized by the class size. Hence, it clearly displays the classifier's confusion between two classes, if exists.
2) **Correct Classification Rate (Accuracy)** is assessed by comparing the classification rate with ground truth labeling of the test set. The accuracy is defined as the number of correct classifications divided by the total number of examples in the test set.
3) **Balanced Error Rate (BER)** is the average of the number of incorrect classifications for each class, normalized by the class size. The BER criteria is especially useful when the number of instances vary highly among different classes.

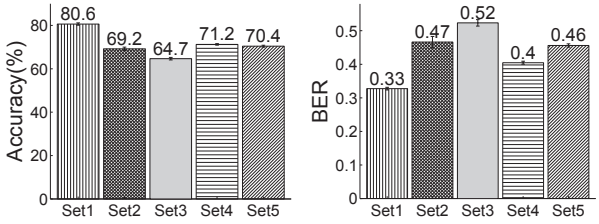
6 Results and Discussion

6.1 Classification Results with Individual Feature Sets

Initially, we investigate the utility of using isolated feature sets for classifying the pattern classes. A separate model is trained with each feature set in Table 1 to discover how well these features capture the significant characteristics of the interaction pattern classes. The recognition performance of training with individual feature sets can be seen in Fig. 12, along with the confusion matrices in Fig. 13.

The classifier trained with Set 1 (agent force-related features) achieves the best classification performance with an accuracy of 80.6% and a BER of 0.33. On the other hand, the classifier trained with Set 3 (interactive force-related features) yields the lowest performance with 64.7% accuracy and BER of 0.52.

Note that even though all classifiers achieve recognition accuracies higher than 60%, the BERs are comparatively



(a) Accuracy (b) Balanced Error Rate

Fig. 12. Classification results of individual feature sets. Set 1: Agent force-related feature set, Set 2: Net force-related feature set, Set 3: Interactive force-related feature set, Set 4: Velocity-related feature set, Set 5: Power-related feature set.

high (≥ 0.3). Examining the confusion matrices in depth (see Fig. 13), we observe that each individual feature set is successful[8] in recognizing at least 4 interaction patterns, but have confusions in one or two classes. Specifically, the classifiers trained individually with Sets 1 and 2 perform poorly in the classification of C3: Agent-force related features in Set 1 suffer from confusion between C3 and C5, whereas net force-related features in Set 2 confuses C3 with both C5 and C6. As seen in Fig. 10, the mean magnitudes of individual forces are close to each other for C3 and C5, and so are the net force magnitudes of C3, C5, and C6. Hence classifiers trained with these features are indeed expected to confuse the patterns, as isolated features are not descriptive on their own for differentiating between these pattern classes. Similarly, it is no surprise for the classifier trained with the interactive-force related features in Set 3 to confuse C2 and C3 with C5 and C6. Finally, a similar case holds also for the Set 4's velocity- and Set 5's power-related features not being able to differentiate between C4 and C6.

6.2 Classification Results with the Combined Feature Set

The approach described above emphasizes the performance of isolated individual feature sets. However, some features can be used in combination to enhance the accuracy of the recognition of interaction patterns. We construct a *combined feature set*, comprising of all of the features in the aforementioned 5 feature sets. Using the combined feature set, we achieve an increased accuracy of 84.2% and

8. We consider a classification to be unsuccessful in case that the correct classification rate is lower than random recognition rate, $1/6$ in our case.

Set1		PREDICTED INSTANCES					
		C1	C2	C3	C4	C5	C6
ACTUAL INSTANCES	C1	**0.94**	0.01	0.01	0.00	0.02	0.02
	C2	0.03	**0.41**	0.02	0.01	0.43	0.10
	C3	0.05	0.03	**0.14**	0.10	0.45	0.23
	C4	0.00	0.00	0.05	**0.80**	0.15	0.00
	C5	0.00	0.03	0.03	0.04	**0.78**	0.12
	C6	0.00	0.00	0.00	0.00	0.03	**0.97**

(a) Agent force-related feature set

Set2		PREDICTED INSTANCES					
		C1	C2	C3	C4	C5	C6
ACTUAL INSTANCES	C1	**0.95**	0.03	0.00	0.00	0.00	0.02
	C2	0.05	**0.45**	0.05	0.00	0.32	0.13
	C3	0.03	0.05	**0.16**	0.00	0.36	0.40
	C4	0.00	0.00	0.00	**0.17**	0.11	0.72
	C5	0.00	0.08	0.03	0.02	**0.60**	0.27
	C6	0.00	0.00	0.00	0.01	0.10	**0.89**

(b) Net force-related feature set

Set3		PREDICTED INSTANCES					
		C1	C2	C3	C4	C5	C6
ACTUAL INSTANCES	C1	**0.75**	0.00	0.00	0.00	0.03	0.22
	C2	0.05	**0.10**	0.03	0.00	0.47	0.35
	C3	0.08	0.01	**0.03**	0.06	0.49	0.33
	C4	0.00	0.01	0.10	**0.55**	0.30	0.04
	C5	0.02	0.06	0.04	0.04	**0.59**	0.25
	C6	0.11	0.01	0.00	0.00	0.02	**0.86**

(c) Interactive force-related feature set

Set4		PREDICTED INSTANCES					
		C1	C2	C3	C4	C5	C6
ACTUAL INSTANCES	C1	**0.88**	0.02	0.01	0.00	0.04	0.05
	C2	0.06	**0.60**	0.02	0.00	0.20	0.12
	C3	0.08	0.02	**0.62**	0.00	0.07	0.21
	C4	0.00	0.00	0.00	**0.00**	0.12	0.88
	C5	0.01	0.02	0.02	0.00	**0.61**	0.34
	C6	0.00	0.01	0.01	0.00	0.11	**0.87**

(d) Velocity-related feature set

Set5		PREDICTED INSTANCES					
		C1	C2	C3	C4	C5	C6
ACTUAL INSTANCES	C1	**0.86**	0.05	0.04	0.00	0.03	0.02
	C2	0.11	**0.25**	0.10	0.00	0.47	0.07
	C3	0.02	0.04	**0.64**	0.00	0.16	0.14
	C4	0.00	0.00	0.02	**0.04**	0.10	0.84
	C5	0.02	0.08	0.05	0.01	**0.58**	0.26
	C6	0.00	0.01	0.01	0.00	0.05	**0.93**

(e) Power-related feature set

Combined Set		PREDICTED INSTANCES					
		C1	C2	C3	C4	C5	C6
ACTUAL INSTANCES	C1	**0.95**	0.00	0.00	0.00	0.02	0.03
	C2	0.03	**0.73**	0.00	0.02	0.14	0.08
	C3	0.00	0.01	**0.70**	0.03	0.14	0.12
	C4	0.00	0.00	0.02	**0.86**	0.08	0.04
	C5	0.00	0.01	0.03	0.05	**0.74**	0.17
	C6	0.00	0.00	0.01	0.00	0.07	**0.92**

(f) Combined feature set

Fig. 13. Confusion matrices of classifiers trained with individual feature sets and the combined set. Highlighted cells indicate significant misclassifications. C1: Harmonious translation, C2: Harmonious rotation with translation, C3: Harmonious braking, C4: Persistent conflict, C5: Jerky conflict, C6: Passive agreement

a reduced BER of 0.19. The reduced BER value illustrates the increased discriminative power of the combined set in inhibiting the misclassifications. The confusion matrix of the classifier trained with the combined feature set is given in Fig. 13(f). Upon closer inspection, we observe that unlike the classifiers trained with individual feature sets, this classifier is able to recognize all of the interaction patterns without significant confusion. Particularly, it achieves the highest improvement for classification of C2 (Harmonious rotation with translation), C3 (Harmonious braking) and C4 (Persistent conflict), all of which had poor recognition performance with individual feature sets.

6.3 Selection of the Optimal Feature Set

The final step in our learning approach is to select the most informative features in the combined feature set. This is motivated by the fact that the combined set gets quite large as a result of aggregating 5 individual feature sets. This manifests itself in a gradual increase in the running time of model selection and training as the datasets get larger. Furthermore, the combined set may contain some unnecessary and even irrelevant features, which may lead to inferior classification performance. Such features should be removed to enhance the recognition accuracy. Hence, we utilize the Minimum Redundancy Maximum Relevance (mRMR) feature selection algorithm to select most promising features [21].

The mRMR algorithm computes k maximally relevant and minimally redundant features from a larger feature space of size K, consisting of all 576 features in our case, where $k = 1, 2, ..., K$. In the end, the feature set that yields the highest accuracy is declared as the optimal feature set for the recognition of interaction patterns. Fig. 14 shows the classification accuracies against the number of features in the set. This diagram illustrates that the optimal feature set consists of 243 features. This optimal set achieves a performance even better than that of the combined feature

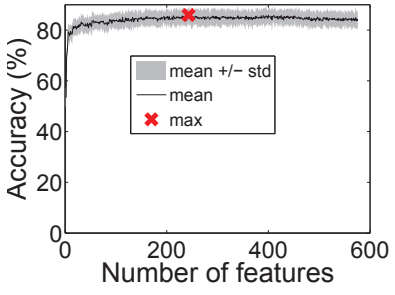


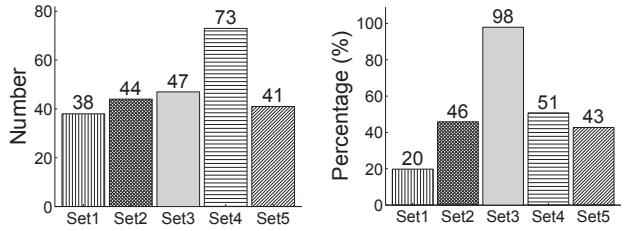
Fig. 14. Classification accuracies for the feature sets, which are built incrementally using mRMR, plotted against the number of features in the features sets. The red cross denotes the optimum feature set, which yields the highest accuracy.

Optimal Set		PREDICTED INSTANCES					
		C1	C2	C3	C4	C5	C6
ACTUAL INSTANCES	C1	**0.94**	0.00	0.01	0.00	0.03	0.02
	C2	0.03	**0.72**	0.02	0.03	0.15	0.05
	C3	0.00	0.01	**0.72**	0.04	0.12	0.11
	C4	0.00	0.00	0.00	**0.85**	0.08	0.07
	C5	0.00	0.02	0.01	0.04	**0.77**	0.16
	C6	0.00	0.00	0.00	0.00	0.04	**0.96**

Fig. 15. Confusion matrix of classifier trained with the optimal feature set. C1: Harmonious translation, C2: Harmonious rotation with translation, C3: Harmonious braking, C4: Persistent conflict, C5: Jerky conflict, C6: Passive agreement

set with an accuracy of 86% and a BER of 0.18. The confusion matrix of the classifier trained with the optimal feature set is given in Fig. 15. We observe that the classifier can successfully recognize all six interaction patterns.

Fig. 16 presents the numbers and the percentages of the features in the optimal feature set taken from individual sets. At first glance, Fig. 16(a) gives the impression that Set 4 (velocity-related features) is a superior feature representation because of its large contribution to the optimal



(a) Number of features. (b) Percentages of features.

Fig. 16. Number and percentage of features from individual feature sets, contributing to the optimal feature set.

feature set; however this is partly due to the high number of features in the initial set. The percentages of features contributed by each individual feature set provides more meaningful information. As demonstrated in Fig. 16(b), almost all of the features in Set 3 (interactive force-related features) eventually contribute to the optimal feature set. On the other hand, almost half of Set 4 is discarded during feature selection.

In this study, we demonstrate that feature sets presented in Section 5.3 are complementary. Moreover, we illustrate the significance of feature selection in accomplishing higher recognition accuracies. As happened in our case, the inclusion of many features may diminish recognition performance unless all are collectively relevant. However, it is worth noting that there is a trade-off between the processing required for optimal feature selection and the resulting gains in the accuracy.

7 Conclusions and Future Work

This work is a first step in discovering patterns in haptic interaction between humans. Specifically, we present a taxonomy of conflict-originated interaction patterns and a method for the classification of these patterns in physical collaboration scenarios, where two humans communicate through the haptic channel. Six interaction patterns were identified based on the interaction of 20 human dyads who transport a virtual object to certain goal positions in a haptics-enabled simulation environment. Time-series interaction data was divided into segments, each of which was labeled by an annotator, who monitors the interaction from outside.We proposed five distinct feature sets, four of which consist of haptic features, to recognize the interaction patterns. We demonstrated that haptic features exhibit significant information about the interaction between partners, and the classifier trained with a combination of haptic and velocity-related features achieves a correct classification rate of 86%.

The proposed taxonomy offers several layers to understand the interaction between partners. We suggest that this taxonomy can be useful for different applications at different levels. For instance, many applications would only be interested in discovering the interaction at a very high level, such as whether agents act harmoniously or not. Specifically, the first two layers of the taxonomy, "Motion Intentions" and "Interaction Types", are general, whereas the last layer, "Interaction Patterns", is more task-dependent. However, the interaction patterns defined in this layer can be modified based on the particular task in hand.

On the other hand, the machine learning approach we present here is generic, and given training data, generalizable to numerous tasks, which involve the interaction of multiple humans and/or robots in both direct or indirect communication. The classification and feature selection ideas we adopt are completely task independent and are usable whenever the behavior labels are defined and descriptive features are extracted from data. However, it should be noted that the features that we have identified as descriptive in our task may not directly apply to other tasks. Hence, more labor should be paid to discovering the most appropriate features in a task-dependent manner.

One shortcoming of our learning approach is its being fully supervised, requiring all interaction data to be intact and labeled before classification. This effectively makes the data collection and annotation stages very time consuming; thus puts a restriction on the amount of data we worked with in this paper. Upon collecting and annotating the data, we obtained a labeled data set consisting of 1944 instances. Even though the data size seems moderate, we would like to express that it was already large and the proposed analysis took significant amount of time.[9] In the future, we intend to apply unsupervised or semi-supervised learning methods for classification to enable online intention prediction during an ongoing collaboration. Our final goal is to develop a robot, which can infer about interaction patterns in real-time and collaborate with its human partner(s) accordingly.

This study reveals that interaction behaviors can be manifested through the forces that the agents apply when interacting with each other. However, we would like to express that these behaviors are also strongly influenced by the characters and emotional states of the individuals. In fact, such characteristics can be discovered through data mining techniques by investigating the way each agent applies forces on the object. For instance, a human who applies large forces on the object might be considered dominant, whereas another who changes his/her forces frequently can be seen as being inept. A future direction aims at discovering such individual characteristics during a collaboration task. Similarly, the findings of this study can be supported through a data-driven approach to build an interaction behavior taxonomy from observations.

References

[1] A. L. Berezm. Review of eudico linguistic annotator (elan). In *Language Documentation and Conservation*, volume 1, pages 283–289, December 2007.

[2] A. Billard, S. Calinon, R. Dillmann, and S. Schaal. Robot programming by demonstration. In B. Siciliano and O. Khatib, editors, *Springer Handbook of Robotics*, pages 1371–1394. Springer, 2008.

9. On an 8 core Intel Core i7-3770 CPU @ 3.40GHz machine with 16 GB memory, the processing time required for the SVM classification with parameter optimization was 65 minutes. Additionally, the mRMR feature selection procedure took around 7.5 days.

[3] A. Bussy, P. Gergondet, A. Kheddar, F. Keith, and A. Crosnier. Proactive behavior of a humanoid robot in a haptic transportation task with a human partner. In *RO-MAN, 2012 IEEE*, pages 962–967. IEEE, 2012.
[4] T. Carlson and Y. Demiris. Collaborative control for a robotic wheelchair: Evaluation of performance, attention, and workload. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 42(3):876–888, 2012.
[5] B. Corteville, E. Aertbeliën, H. Bruyninckx, J. D. Schutter, and H. V. Brussel. Human-inspired robot assistant for fast point-to-point movements. In *IEEE International Conference on Robotics and Automation, ICRA*, pages 3639–3644, 2007.
[6] V. Duchaine and C. M. Gosselin. General model of human-robot cooperation using a novel velocity based variable impedance control. In *Second Joint EuroHaptics Conference, 2007 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pages 446–451. IEEE, 2007.
[7] P. Evrard, E. Gribovskaya, S. Calinon, A. Billard, and A. Kheddar. Teaching physical collaborative tasks: object-lifting case study with a humanoid. In *Humanoid Robots, 2009. Humanoids 2009. 9th IEEE-RAS International Conference on*, pages 399–404. IEEE, 2009.
[8] P. Evrard and A. Kheddar. Homotopy switching model for dyad haptic interaction in physical collaborative tasks. In *WHC'09: IEEE World Haptics Conference*, pages 45–50, 2009.
[9] T. Flash and N. Hogan. The coordination of arm movements: an experimentally confirmed mathematical model. *The Journal of Neuroscience*, 5(7):1688–1703, 1985.
[10] R. Groten, D. Feth, H. Goshy, A. Peer, D. Kenny, and M. Buss. Experimental analysis of dominance in haptic collaboration. In *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, pages 723–729, 2009.
[11] R. Groten, D. Feth, R. Klatzky, and A. Peer. The role of haptic feedback for the integration of intentions in shared task execution. *IEEE Transactions on Haptics*, 6(1):94–105, 2013.
[12] N. Jarrassé, T. Charalambous, and E. Burdet. A framework to describe, analyze and generate interactive motor behaviors. *PLOS One*, 7(11):e49945, 2012.
[13] A. Kheddar. Human-robot haptic joint actions is an equal control-sharing approach possible? In *Human System Interactions (HSI), 2011 4th International Conference on*, pages 268–273, 2011.
[14] K. Krippendorff. Reliability in content analysis. *Human Communication Research*, 30(3):411–433, 2004.
[15] A. Kucukyilmaz, T. Sezgin, and C. Basdogan. Conveying intentions through haptics in human-computer collaboration. In *WHC'11: IEEE World Haptics Conference*, pages 421–426, June 2011.
[16] A. Kucukyilmaz, T. Sezgin, and C. Basdogan. Intention recognition for dynamic role exchange in haptic collaboration. *Haptics, IEEE Transactions on*, 6(1):58–68, 2013.
[17] M. Lawitzky, A. Mörtl, and S. Hirche. Load sharing in human-robot cooperative manipulation. *Proc. of IEEE Int. Symposium in Robot and Human Interactive Communication*, pages 185 –191, 2010.
[18] A. Melendez-Calderon, V. Komisar, G. Ganesh, and E. Burdet. Classification of strategies for disturbance attenuation in human-human collaborative tasks. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, pages 2364–2367. IEEE, 2011.
[19] A. Mörtl, M. Lawitzky, A. Kucukyilmaz, T. M. Sezgin, C. Basdogan, and S. Hirche. The role of roles: Physical cooperation between humans and robots. *Int. J. Robotic Res.*, 31(13):1656–1674, 2012.
[20] S. Oguz, A. Kucukyilmaz, T. Sezgin, and C. Basdogan. Haptic negotiation and role exchange for collaboration in virtual environments. In *Haptics Symposium, 2010 IEEE*, pages 371–378, 2010.
[21] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(8):1226–1238, 2005.
[22] M. Rahman, R. Ikeura, and K. Mizutani. Control characteristics of two humans in cooperative task and its application to robot control. In *Industrial Electronics Society, 2000. IECON 2000. 26th Annual Confjerence of the IEEE*, volume 3, pages 1773–1778. IEEE, 2000.
[23] K. B. Reed and M. A. Peshkin. Physical collaboration of human-human and human-robot teams. *IEEE Trans. Haptics*, 1(2):108–120, 2008.
[24] O. Schrempf, U. Hanebeck, A. Schmid, and H. Woern. A novel approach to proactive human-robot cooperation. In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, pages 555–560, 2005.
[25] N. Stefanov, A. Peer, and M. Buss. Role determination in human-human interaction. In *WHC'09: IEEE World Haptics Conference*, pages 51–56, Salt Lake City, USA, 2009.
[26] T. Takeda, Y. Hirata, and K. Kosuge. Dance step estimation method based on HMM for dance partner robot. *IEEE Transactions on Industrial Electronics*, 54(2):699 – 706, 2007.
[27] T. Tsumugiwa, R. Yokogawa, and K. Hara. Variable impedance control with virtual stiffness for human-robot cooperative peg-in-hole task. In *Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference on*, volume 2, pages 1075–1081, 2002.
[28] Z. Wang, A. Peer, and M. Buss. An hmm approach to realistic haptic human-robot interaction. In *WHC'09: IEEE World Haptics Conference*, pages 374–379, Salt Lake City, USA, March 2009.
[29] J. E. Young, J. Sung, A. Voida, E. Sharlin, T. Igarashi, H. I. Christensen, and R. E. Grinter. Evaluating human-robot interaction. *International Journal of Social Robotics*, 3(1):53–67, 2011.

Cigil Ece Madan received her BS and MS degrees in Mechanical Engineering from Koc University, Turkey. Her research interests include human-X haptic interaction in virtual domains and haptic user interfaces with a focus on human-robot interaction.

Ayse Kucukyilmaz is currently a research associate at the Personal Robotics Laboratory of Imperial College London. She received her PhD degree in Computational Sciences and Engineering from Koc University in 2013. She obtained her BS and MS degrees in Computer Engineering from Bilkent University, respectively in 2004 and 2007. She Her research interests include haptics, physical human-X interaction, assistive robotics, adaptive user modeling, affective computing, social robotics, and machine learning.

Tevfik Metin Sezgin graduated summa-cum-laude with Honors from Syracuse University in 1999. He received his MS and PhD degrees from Massachusetts Institute of Technology in 2001 and 2006. He subsequently joined the University of Cambridge as a Postdoctoral Research Associate, and held a visiting researcher position at Harvard University in 2010. He is currently an assistant professor at Koc University, Istanbul. His research interests include intelligent human-computer interfaces and HCI applications of machine learning. He is particularly interested in applications of these technologies in building intelligent pen-based interfaces.

Cagatay Basdogan is a faculty member in the Mechanical Engineering and Computational Sciences and Engineering programs of Koc University, Istanbul, Turkey. He is also the director of the Robotics and Mechatronics Laboratory at Koc University. Before joining Koc University, he worked at NASA-JPL/Caltech, MIT, and Northwestern University Research Park. His research interests include haptic interfaces, robotics, mechatronics, biomechanics, medical simulation, computer graphics, and multi-modal virtual environments. Prof. Basdogan received his PhD degree in Mechanical Engineering from Southern Methodist University in 1994. He is currently the associate editor in chief of IEEE Transactions on Haptics and the associate editor of Computer Animation and Virtual Worlds journals.