Online Identification of Interaction Behaviors from Haptic Data during Collaborative Object Transfer

Ayse Kucukyilmaz[1], Illimar Issak[1]

Abstract—Joint object transfer is a complex task, which is less structured and less specific than what is existing in several industrial settings. When two humans are involved in such a task, they cooperate through different modalities to understand the interaction states during operation and mutually adapt to one another's actions. Mutual adaptation implies that both partners can identify how well they collaborate (i.e. infer about the interaction state) and act accordingly. These interaction states can define whether the partners work in harmony, face conflicts, or remain passive during interaction. Understanding how two humans work together during physical interactions is important when exploring the ways a robotic assistant should operate under similar settings. This study acts as a first step to implement an automatic classification mechanism during ongoing collaboration to identify the interaction state during object co-manipulation. The classification is done on a dataset consisting of data from 40 subjects, who are partnered to form 20 dyads. The dyads experiment in a physical human-human interaction (pHHI) scenario to move an object in an haptics-enabled virtual environment to reach predefined goal configurations. In this study, we propose a sliding-window approach for feature extraction and demonstrate the online classification methodology to identify interaction patterns. We evaluate our approach using 1) a support vector machine classifier (SVMc) and 2) a Gaussian Process classifier (GPc) for multi-class classification, and achieve over 80% accuracy with both classifiers when identifying general interaction types.

Index Terms—Classification, Feature Extraction, Force and Tactile Sensing, Haptics and Haptic Interfaces, Human Factors and Human-in-the-Loop, Learning and Adaptive Systems, Physical Human-Human Interaction, Physical Human-Robot Interaction, Recognition

I. INTRODUCTION

HUMANS collaborate through tiresome and demanding physical activities during their daily routines, such as lifting and moving objects. With recent developments in the field, humanoid robots show great potential as helpers, which can proactively aid humans in a wide range of physical collaboration scenarios. However, even though these tasks seem trivial to us, the variability in human operations impose a big challenge for robots.

Human collaboration is complex; it involves strong interpersonal coordination and mutual role adaptation mechanisms [1]. These help humans to determine how and when their partner's goals and the overall interaction states change, allowing them
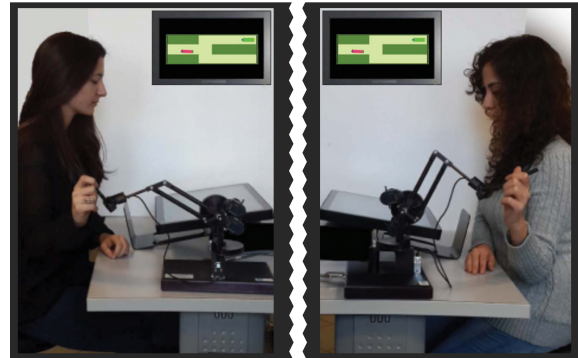
Fig. 1. A dyad using haptic interfaces to collaboratively transfer an object to a goal configuration in a virtual environment, shown in Fig. 2.

to enhance their movements [2]. Determining how and when interaction behaviors change is a key issue in identifying the underlying sensorimotor processes for collaboration and assistance during physical human-human interaction (pHHI). This fundamental understanding also has implications for physical human-robot interaction (pHRI). A robot, which can instantaneously infer about how well it interacts with a human would be able to better complement the human operation as an assistant. Such proactivity would facilitate the design of many pHRI applications, ranging from tele-operation to robot-assisted surgery and rehabilitation.

This study explores how two human partners' interactive states change over physical collaboration. We use the data collected over a dyadic object transfer task, generated using the setup shown in Figure 1. Using the interaction behavior taxonomy proposed in [3], we investigate an online feature extraction method and perform online classification for distinguishing between interaction states during pHHI. This is a first step to build a proactive robotic partner, which can assist a human, while being aware of the interaction state that the partners are in.

II. BACKGROUND

Even though HHI has been studied by many researchers to transfer physical collaboration skills to human-robot teams [4], [5], [6], [7], [8], [9], not many studies inform the decision process based on an understanding of human-human interaction (HHI). In fact, there are only a few studies that made an effort to define interaction patterns for pHHI. Melendez-Calderon et al. [10] used interaction forces and muscle activation to classify specialization strategies in a tracking task. The classification was rule-based and the framework was not robust against the addition of new interaction strategies. Defined roles were static and rather than attempting to distinguish between

these, the study discussed their efficacy for specific tasks. In this sense, this study did not attempt to provide a general classification of the interactive behaviors.

One of the most important studies for formalising human-human interaction is by Jarrassé et al. [11], who proposed a taxonomy, which identified several types of role arbitration in shared control based on game theory and neuroscience concepts. The interactive behaviors were classified as *competition*, *collaboration* and *cooperation*, and made use of cost functions that minimize error and effort to distinguish between these classes. Among these types, we are mostly interested in *collaborative* arbitration policies, which are most suitable for general co-manipulation tasks.

In the context of pHRI, shared control is a way to take into account human operation characteristics to proactively alter robotic operation. This may involve enabling the human-robot team to dynamically arbitrate leader and follower roles as done in [12], [13], [14], [15], [16], [17], [18], [19]. Collaborative arbitration is defined by dynamic role exchanges between the human and the robot. Powell and O'Malley [20] implemented different roles for creating different motor interactions and investigated the resulting physical interactions in shared control. Oguz et al. [21] programmed robotic behaviors based on game theory in a two-party physical negotiation scenario. However, even though these roles might be considered human-inspired, none of these studies were informed by observations from real pHHI data.

Our study aims to distinguish between different collaborative interaction states during pHHI with an aim to create a computational model for physical human collaboration. An interaction state may point to different aspects of collaboration, such as joint attention, fatigue, role hierarchy, partner specialization behaviors, etc. In our context, we focus on the harmony between the collaborating agents as a quality of partnership to define interaction states. In our earlier work [3], we proposed a taxonomy of interactive behaviors, defining three main classes of collaborative interaction types, namely harmonious, conflicting and neutral behaviors. This is a general and universal taxonomy that provides an interaction-oriented characterization for any physical collaborative task. In comparison to [11], ours provides a simpler characterization, defining three universal classes of interaction types. Also, being a hierarchical taxonomy, it is able to capture different interaction patterns specifically observed in a given task. In [3] six task-dependent interaction patterns were identified and these patterns were automatically classified through supervised learning. The study identified some descriptive features of interaction; however, it required extensive human annotation, and classification was done in an offline fashion. The current study builds on this earlier work [3] and improves it by introducing an online feature extraction technique, which enables classification to be done during ongoing collaboration.

This online classification approach requires working on timeseries data. In robotics research, timeseries analysis has been performed to estimate intentions and recognize actions. Lin et al. [22] employed a sliding window approach over the raw timeseries to extract features for distinguishing between expert and intermediate surgeons in a 4-throw suturing task

using the da Vinci robotic surgical system. They used the concatenation of raw data measurements enclosed by the time window and merged the data from separate channels of information to form a super-feature vector. Stefanov et al. [23] adopted a similar approach for distinguishing between the transportation and positioning phases of point-to-point movement in a haptics enabled virtual environment. However, instead of concatenating raw timeseries data, they concatenated features extracted from timeseries data as proposed by Olszewski [24]. This method performed an amplitude discretization of the continuous data by defining several thresholds to split the signal into classes according to the signal magnitude. This discretization resulted with reducing the signal to a binary number, which served as the codebook feature vector. In this study, we investigate the use of descriptive statistics of raw data as features.

This study proposes an online feature extraction method and illustrate its use for timeseries classification on a dataset generated in [3]. The data is recorded as two humans remotely collaborate through haptic interfaces to transport an object in a virtual environment, and labelled according to Madan et al.'s interaction behavior taxonomy [3]. In order to enable online classification on this data, we use a sliding window approach for feature extraction. The proposed approach uses a fixed-size window, over which a constant number of features are computed using the multidimensional timeseries data within the window. This window is then shifted forwards by a fixed amount as time advances, while extracting features on the way. The classification is done for each window, and by setting the window size small enough, we enable online estimations for interaction states.

III. HHI INTERACTION DATASET

This study builds on our earlier work and uses the interaction behavior taxonomy proposed in [3] to investigate an online feature extraction and classification technique for identifying interaction states on-the-fly during pHHI. We analyze the data collected in a virtual environment, where two humans interact through the haptic channel. The dataset used in this study is publicly available at https://github.com/aysekyz/HHIBehaviorDataset/.

A. HHI Experiment

40 volunteers (9 female and 31 male, aged between 21 and 29, all right-handed) participated in the study. Subjects were randomly matched to form dyads. Each partner within a dyad sat in a separate room in front of a computer screen, which displayed a maze-like virtual scene. Subjects interacted with a Sensable (currently Geomagic) Phantom Premium haptic device using a stylus attachment as in Figure 1. The haptic devices were connected to separate PCs and communicated through a dedicated UDP connection over the local network.

The subjects were asked to coordinate their motion and jointly move an object to match a target configuration. In order to provoke a range of different interaction patterns, two different scenes, named straight and bifurcated, were created as shown in Figure 2. The scenes were designed so that it was
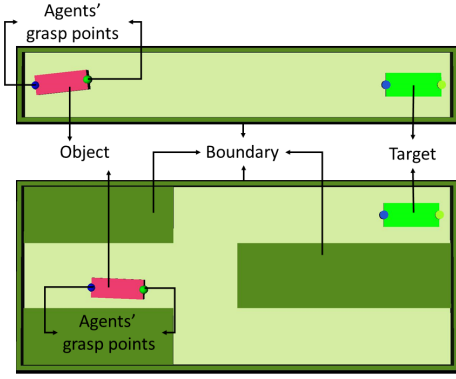
Fig. 2. The subjects were shown two scenes in the virtual environment as part of the experiment. The straight scene (top) and the bifurcated scene (bottom). The object (pink rectangular prism), target (bright green rectangular area), boundaries (dark green walls), and the agents' grasping points (blue and green spheres on the object) are shown.

possible to observe interaction patterns during rotation and translation of the object. In each scene, five different manipulation scenarios were created[1]. Each manipulation scenario was presented twice in both straight and bifurcated scenes, ensuring that each dyad has experimented with each scenario exactly twice in the course of 10 trials. The order of the scenarios was the same between the first and the second set of 5 trials. As a result of this procedure, 400 interaction sequences (2 scenes × 20 dyads × 10 trials) are recorded.

At the beginning of the experiments, the dyads were presented with two practice trials in order to familiarize them with the setup and the task. In order to balance the learning effects, dyads were assigned random numbers and the order of the scenarios were permuted using a Latin square design depending on this number. 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. The purpose of the task was to complete the object transfer task, parking the object at a target configuration and staying there for a predetermined period of 5 seconds before moving on to another goal. During the experiment, the subjects were presented with two different scenes to observe interaction patterns in both translational and rotational motion. The first scene, dubbed the *straight scene*, depicts a horizontal path, whereas the second scene, called the *bifurcated scene*, presents a fork-shaped path for the users to follow.

The experiment involved using visual and haptic cues when completing the task. The users were given proprioceptive information by showing them their grasping points, represented as blue and green spheres attached to the object. In each trial, a single target was depicted with a green rectangle with clear

orientation information. The target changed color to blue once the object reached the desired configuration, and a countdown timer was shown in the middle of the screen, informing the dyad of the time they needed to wait before moving onto another target configuration. If the dyad managed to wait in the desired configuration for 5 seconds, a new target appeared, starting a new trial.

The manipulated object was physically modeled as a rigid body of 0.4 kg, which could move in 2D, in response to forces applied on it. The translational and rotational inertia was computed using the mass. The movement of the haptic interfaces were used to compute the interaction forces applied on the object. In particular, the end-effector positions of the haptic styli along x- and z-axes mapped to the positions of the individual *haptic interface points* (HIPs) in the virtual world. The movement was constrained to a virtual plane parallel to the ground, where motion in x- and z-axes respectively controlled right/left and forward/backward directions with respect to where the user sat. We programmed two spring-damper systems between each agent's HIP and their grasping point on the object to compute the individual forces $\boldsymbol{F_{HIP_1}}$ and $\boldsymbol{F_{HIP_2}}$ applied by the agents on the object as follows:

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

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

where $K_p = 0.25$ N/mm and $K_d = 0.001$ Ns/mm are spring and damper coefficients, $\boldsymbol{x_{HIP_1}}, \boldsymbol{x_{HIP_2}}, \boldsymbol{\dot{x}_{HIP_1}}, \boldsymbol{\dot{x}_{HIP_2}}$ are the positions and velocities of HIPs, and $\boldsymbol{x_{g_1}}, \boldsymbol{x_{g_2}}, \boldsymbol{\dot{x}_{g_1}}, \boldsymbol{\dot{x}_{g_2}}$ are the positions and velocities of the grasping points of the agents. The agents were reciprocally fed back with forces $-\boldsymbol{F_{HIP_1}}$ and $-\boldsymbol{F_{HIP_2}}$ through the haptic devices, so that they could feel the dynamics of the object[2].

In case of collisions with the boundaries, an impact force $\boldsymbol{F_I}$ and an impact moment M_I were computed based on the penetration depth of the object into the boundaries using the calculated object pose. This calculated force vector was mapped over the object, assuming 2D dynamics. Also, translational and rotational friction ($\boldsymbol{F_f}$ and M_f), calculated using the Coulomb friction model were applied on the object to affect its motion[3].

B. Interaction Behavior Taxonomy

The taxonomy assumes that there are three main types of interaction in any collaborative task between humans: 1. working in harmony, 2. coping with conflicts, 3. remaining passive during interaction. With this understanding, frequently observed patterns, which were specific to this task domain, were defined and a taxonomy of physical interaction was created as shown in Figure 3. The taxonomy includes 6 task-dependent interaction pattern classes, which fall into 3 task-independent interaction types:

---

[1]Please note that [3] defines four interaction scenarios, in which conflicts between partners are artificially invoked by providing each agent with different visual information about the location of the target configuration. In one of the scenarios, only one subject is shown a goal, making the other agent blind to the goal. In order to balance the experience, this scenario is repeated twice, so that both partners act as the blinded subject at some point during the experiment. For brevity, here, we say five different scenarios were created in [3], considering the two dual sub-scenarios as separate.

[2]Due to mechanical constraints of the haptic devices, the forces fed back to the humans were thresholded at 4.0 N.

[3]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. This created a slippery dynamic environment, requiring good communication for fine control.
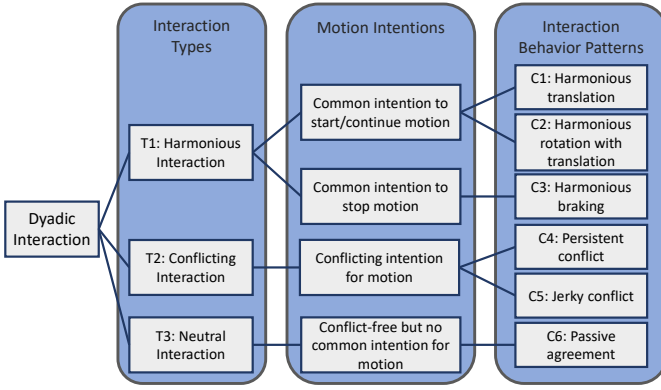
Fig. 3. Taxonomy of physical interactions in dyadic object manipulation.

T1: Harmonious interaction: C1: *Harmonious Translation*; C2: *Harmonious Rotation with Translation*; C3: *Harmonious Braking*

T2: Coping with conflicts: C4: *Persistent Conflict*; C5 *Jerky Conflict*

T3: Neutral interaction: C6: *Passive Agreement*

The data collected in the experiment was manually labelled following the taxonomy definitions by two independent annotators, resulting with a populated dataset of 1944 instances of different lengths. This is equivalent to more than 6 hours-worth of physical interaction data. The details of the annotation process and the dataset can be found in [3].

IV. METHODOLOGY

The dataset presented in the previous section contains variable length interaction behaviors segments, which are labeled to belong to a single interaction class. In order to perform online identification of pHHI behaviors, we adopt a sliding-window approach to capture smaller chunks of potentially meaningful data sequences during interaction. We set fixed-sized windows over the whole interaction data and compute features over each window. Using these features, we perform classification and compare the predictions to the ground truth annotations. Even though this study presents these results over recorded data, the approach can be transferred to real-time interaction once classification models are trained.

In the rest of this section, we describe our end-to-end methodology: As a first step, we prepare the dataset for classification and perform feature extraction over sliding windows. We test our features with two different classifiers, namely a support vector machine (SVMc) and a Gaussian process (GPc) classifier. Since we use sliding windows on annotated data, our approach is prone to capture interaction segments which cover data belonging to more than one class. Systematic post-processing is performed to conservatively select a matching label for such interaction chunks, which can then be matched with the ground truth to compute the classification accuracy.

A. Online Feature Extraction

Prior to starting with online classification, we prepare the raw timeseries data for processing. This is done by assigning every data point a label, which matches the annotation defining the class of the corresponding interaction segment. Using this labeled timeseries data, we set a short window to extract a small time sequence to be used for feature extraction. In the experiments, we used window size as 2s, and sliding distance as 1.5s. In other words, we extract 2s worth of features from the interaction every 1.5s. These parameters are empirically set to respect human perception-response time for physical actions and to ensure that each iteration mines enough data for accurate classification. [25] indicates that the perception of an active force cue can take around 1 seconds. This, added with the kinesthetic reaction time for responding to the force cue, would render 2 seconds as a reasonable choice to mimic online collaboration behavior with a human.

As we iterate through the interactions, we hit some windows that contain segments belonging to more than one class. For example when we are moving the window halfway over Class 1, the other half of the window may cover some Class 2 interaction data. In these cases, there is an ambiguity about the correct class of that window. When generating the training set, we discard such ambiguous windows in order to keep the model as clean as possible as they would reduce the performance of the classifier. For this purpose, we check whether the difference between the most prominent label and second most prominent label is less than 20%. If that is true, we drop the window, else we extract the features as usual and add the most prominent label to the match this instance in the training set. With this method, approximately 8% of the windows were dropped. This means that on average, every 24 seconds we encounter some interaction that is too ambiguous to classify as belonging to any of the interaction behaviors.

On the other hand, when testing, discarding such data is neither meaningful nor realistic. When performing online feature extraction in a real world scenario on unseen data, we would have no information about the labels, hence it would be impossible to drop such windows. Hence, for the purposes of our evaluation in this study, we treat the window as belonging to either one of these classes, and we accept any of the candidate classes as a correct classification in case of ambiguity. We consider a window ambiguous if the most occurring class within the window is less than 15% more prominent than the second most occurring label in that window. This is a bit less conservative than our approach for the training set, but allows us more flexibility when identifying possible patterns. Algorithm 1 illustrates the feature extraction process over sliding windows.

Once all labels are set, feature extraction is done within each window. Using the feature definitions in [3], for each window, we compute the mean, standard deviation, median, and interquartile range for each of the variables summarized in Table I. The feature set contains 48 features, which are normalized before being used for training and testing.

B. Model Training

During the experiment each subject conducted 10 trials for each of the two scenes, where the first 5 trials consist of the five scenarios and the second 5 trials repeat the 5 scenarios in

Algorithm 1 Dataset creation over sliding windows

> **procedure** CREATE DATASETS($segments, labels$)
> $X \leftarrow [nil, nil]$ ▷ Create dataset
> $start \leftarrow 0$
> $end \leftarrow start + 2000$ ▷ window size is 2 seconds
> $slider \leftarrow 1500$ ▷ each slide is 1.5 seconds
>   **while** $end < segments.length$ **do**
> $window \leftarrow segments(start : end)$
> $windowLabel \leftarrow labels(start : end)$
> $features \leftarrow ExtractFeatures(window)$
> $mc \leftarrow$ % of most occurring label in window
> $smc \leftarrow$ % of second most occurring label
>   **if** $isTestSet(window) = True$ **then**
>   **if** $mc - smc < 0.15$ **then**   ▷ check ambiguity
> $X \leftarrow [features, mc, smc]$ ▷ add both
>   **else**   ▷ add only most prominent label
> $X \leftarrow [features, mc]$
>   **end if**
>   **else**   ▷ training set
>   **if** $mc - smc < 0.2$ **then** ▷ ambiguous window
>   **continue**   ▷ skip
>   **else**
> $X \leftarrow [features, mc]$
>   **end if**
>   **end if**
> $start \leftarrow start + slider$
> $end \leftarrow end + slider$
>   **end while**
> **end procedure**

TABLE I
FEATURE DEFINITIONS AND EXTRACTED VARIABLES

Variables	# of vars
Agent interaction forces: $\boldsymbol{F_{HIP_1}}, \boldsymbol{F_{HIP_2}}$	4
Net force acting on object: $\boldsymbol{F_{net}} = \boldsymbol{F_{HIP_1}} + \boldsymbol{F_{HIP_2}}$	2
Interactive force acting on object [26]: $$f_i = \begin{cases} \boldsymbol{F_{HIP_{1x}}} & \text{sign}(\boldsymbol{F_{HIP_{1x}}}) \neq \text{sign}(\boldsymbol{F_{HIP_{2x}}}) \\ & \wedge |\boldsymbol{F_{HIP_{1x}}}| \leq |\boldsymbol{F_{HIP_{2x}}}| \\ -\boldsymbol{F_{HIP_{2x}}} & \text{sign}(\boldsymbol{F_{HIP_{1x}}}) \neq \text{sign}(\boldsymbol{F_{HIP_{2x}}}) \\ & \wedge |\boldsymbol{F_{HIP_{1x}}}| > |\boldsymbol{F_{HIP_{2x}}}| \\ 0 & \text{sign}(\boldsymbol{F_{HIP_{1x}}}) = \text{sign}(\boldsymbol{F_{HIP_{2x}}}) \end{cases}$$	1
Object velocity: $\dot{\boldsymbol{x}}_{obj}, \dot{\theta}_{obj}$	3
Power: $P_{HIP_1} = \left| \boldsymbol{F_{HIP_u}} \cdot \dot{\boldsymbol{x}}_{obj} \right| + \left| M_{HIP_u} \dot{\theta}_{obj} \right|$ $P_{HIP_2} = \left| \boldsymbol{F_{HIP_u}} \cdot \dot{\boldsymbol{x}}_{obj} \right| + \left| M_{HIP_u} \dot{\theta}_{obj} \right|$	2

the same order. Considering that the first 5 trials consist of all the scenarios, we determined that a reasonable way to partition the dataset into training and test sets is to use the first 5 trials for training and the remaining 5 for testing. This allows the model to be trained in all interaction scenarios, which would be encountered during testing. In the experiments, data from all dyads are used to train the classifier.

We compared the performances of SVMc and GPc for online classification. For SVMc, we used the scikit-learn implementation. A radial basis function (RBF) kernel is adopted and parameters are optimized by performing a grid search with 5-fold cross-validation. GPy is used as the GPc implementation. The GPc is trained with an RBF kernel, using Bernoulli likelihood and Expectation propagation (EP) inference [27]. Parameter optimization is done by using built-in functionality in GPy, which optimizes the lengthscale and the variance of the kernel against the approximate marginal likelihood. In order to enable multi-class classification, we implemented a one-vs-all scheme for both classifiers.

C. Model Evaluation

We evaluate the performance of our models using confusion matrices and by reporting the correct classification rates. A confusion matrix displays the prediction counts for all classes where diagonal elements represent the number of correctly classified instances and off-diagonal elements represent false positives and false negatives. We normalize the confusion matrix by the number of instances per class to represent the class predictions as percentages. The accuracy of the classifier denote the percentage of correctly classified instances. To get the accuracy we compare all the predicted labels to all the corresponding ground truth labels in the test labels set. If the labels match, then the prediction is set as correct, then the correct classifications are divided by the total number of test labels to get the accuracy. In case of ambiguous labels, we register a correct classification in case the predicted labels match either of the possible labels.

V. RESULTS

We investigate online classification performance of SVMc and GPc in two stages of Madan et al.'s hierarchy [3], on both task-dependent and task-independent behaviors. In doing so, we present two experiments: In the first experiment, we look at how our models perform in distinguishing specific interaction patterns (C1-C6). In the second experiment, we investigate how our approach performs in distinguishing more general behaviors: Harmonious interaction (T1: merged from C1, C2 and C3), Conflicting interaction (T2: merged from C4 and C5) and Neutral interaction (T3: C6). A third experiment is presented to compare the classifier performance for the baseline offline classification approach implemented in [3].

Experiment 1: Online classification of interaction patterns

This experiment investigates the online classification performance for pHHI patterns consisting of six classes (C1 to C6). Our results indicate that SVMc reaches a 78.04% accuracy, whereas GPc achieves an accuracy of 80.79% on the online feature set, with 2.75% improvement on the performance of SVMc. Both classifiers perform extremely well in classifying passive agreement patterns (C6) with an impressive accuracy over 90%.

As can be seen in Figure 4, even though both classifiers work better than a baseline classifier with random guessing, they end up with a strong confusion in recognising C2 and C3. As seen on the confusion matrices, harmonious braking (C3) is often classified as passive agreement (C6). This illustrates that detecting harmonious braking behaviors is harder in online
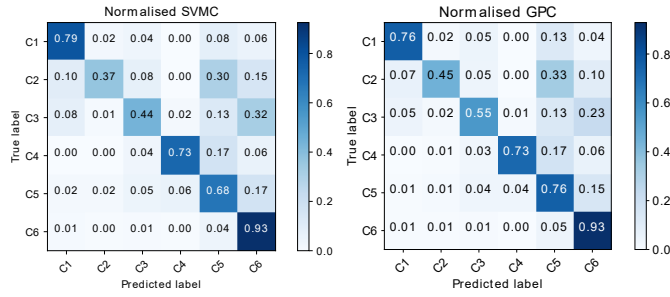
Fig. 4. Confusion matrices for SVMc and GPc for the online classification of interaction patterns

Fig. 5. Confusion matrices for SVMc and GPc for the online classification of interaction types

classification, due to the fact that we are looking at smaller segments, where we do not get to see enough of the interaction to find braking patterns. As observed in [3], the breaking behavior is a complex one: it can be characterized to look similar to harmonious translation at the beginning, followed by some form of conflict and finally a passive agreement. This complex nature of the behavior is hard to capture with our small-sized windows.

Another confusion is observed between harmonious rotation with translation (C2) and jerky conflict (C5). We believe that this, again, might be due the small window-sizes. Jerky conflict typically causes the object to rotate involuntarily or follow undesired trajectories, leading to much lower linear velocity for the object during the course of the interaction pattern. In [3], we had observed that using power or force-related features could not solve the confusion between C2 and C5. On the other hand, the velocity features were very successful in distinguishing between these classes. Since both classes involve rotational forces and velocities, the linear velocity over the course of the whole interaction segment could be deterministic of the correct class. However, the small window sizes we use may not be able to capture such features well enough, leading to lower classification accuracy.

Experiment 2: Online classification of interaction types

The second experiment focuses on identifying general interaction types during pHHI. For this purpose we merge the interaction patterns into pHHI interaction types, to assess the online classification performances on high-level task-independent behaviors. In particular, the interaction patterns are merged into harmonious interaction (T1), conflicting interaction (T2), and neutral interaction (T3), as described in the taxonomy, prior to training the models.

The results indicate that SVMc reaches an 83.31% accuracy when distinguishing between interaction types. This is a 5.27% improvement on the classification performance for interaction patterns. Similarly, GPc achieves an 83.40% accuracy with a 2.61% improvement over the performance of GPc, which was trained to classify the interaction patterns. As seen in Figure 5, in contrast with SVMc, GPc performs a bit better when classifying the neutral interaction types (T3), but the difference is too small to draw any conclusions.
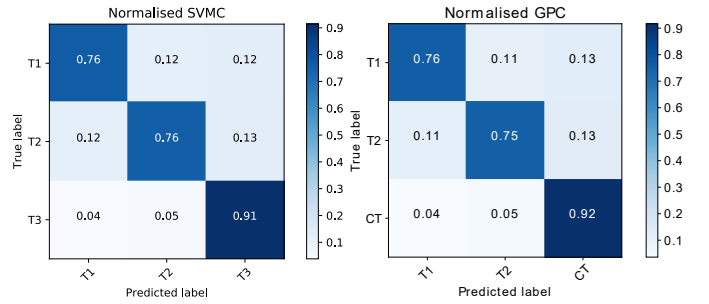
Experiment 3: Offline classification performance

In [3], we adopted similar feature definitions, however we extracted these from within the annotated interaction sequences in the dataset. Each interaction sequence was variable-length, hence we followed a systematic subdivision approach to extract Haar-like features, by dividing the whole interaction segment into support regions and then computing the mean, standard deviation, median, and interquartile range values for each variable within each region. A one-against-one strategy was used with SVMc to obtain classification results through an offline analysis. We call this an offline analysis, because this approach, unlike what we present in the current study, requires a prior knowledge of the length and boundaries of the annotated segments. In this experiment, we verify that we can reproduce the previous results with SVMc and use GPc to get an understanding for the baseline performance, where we have perfect knowledge of behavior boundaries. Table II shows accuracies achieved for offline classification of interaction patterns, which reproduced some of the results presented in [3]. It also presents a summary of the the classification results of the previous two experiments. [3] had reached a classification accuracy of 86% with SVMc, which is consistent with our results. We observe that the use of GPc provides comparable performance with only 1.2% improvement over SVMc for the offline classification.

TABLE II
A COMPARISON OF SVMc AND GPc CLASSIFICATION PERFORMANCES

	Offline classification	Online classification	
	Int. Patterns	Int. Patterns	Int. Types
SVMc	86.1%	78.0%	83.3%
GPc	87.2%	80.8%	83.4%

VI. DISCUSSION AND FUTURE WORK

Our experiments indicate that a combination of haptic data and object velocities can be used to accurately classify human interaction types and patterns in real-time. We also demonstrate an online feature extraction method for timeseries classification to identify interaction states during ongoing collaboration.

Based on the results, we observe that both GPc and SVMc perform well at online classification of interaction states using our feature extraction technique, with GPc achieving slightly

better performance. However, given the training time requirements of GPc, reaching $O(N^3)$, this study demonstrates the feasibility of SVMc to reach good identification performance with much less training time. To be more specific, on an 8-Core Intel i7-4770 CPU @ 3.40GHz machine with 16 GB RAM, learning the SVMc model for classifying interaction patterns with 4992 training samples took approximately 7 minutes. On the other hand, training the GPc with the same number of training instances took approximately 39 hours. Please note that once training is done, the time required for the prediction of the interaction states is negligible and would support real-time interaction.

Our work focused on developing an online classification technique for identifying pHHI behaviors. In physical interaction, haptics plays an important role, which captures a dimension that cannot be represented through other modalities. Despite its importance, the number of studies investigating the physical interaction between the partners and in particular the ones on haptic communication are limited, and the interaction in such systems is still artificial when compared to natural human-human collaboration. In future work, we will design and experiment with sophisticated haptic features, which could provide insight into how much information can be carried over haptics in physical collaboration. We will evaluate the descriptive power of haptic features in determining interaction patterns through feature selection; and we will investigate alternative methods for incorporating time-dependent characteristics in feature definitions to inform the decision process to distinguish between commonly confused classes.

This study presented an evaluation of pHHI behaviors. This approach will be useful to inform the collaboration process when a human works along with a robot as well. As future work, we plan to incorporate interaction state classification in physical human-robot interaction, and program the robot to reactively act in response to the predicted behaviors.

We do not anticipate that our classification method could be trained on one task and tested on another. In this sense, transfer learning was out of the scope of this paper. However, in a future study, we plan to elaborate on how well the trained models respond to changes in task dynamics, as an effort to evaluate the generality of the approach.

REFERENCES

[1] A. Melendez-Calderon, V. Komisar, and E. Burdet, "Interpersonal strategies for disturbance attenuation during a rhythmic joint motor action," *Physiology & behavior*, vol. 147, pp. 348–358, 2015.

[2] A. Takagi, G. Ganesh, T. Yoshioka, M. Kawato, and E. Burdet, "Physically interacting individuals estimate the partner's goal to enhance their movements," *Nature Human Behaviour*, vol. 1, no. 3, p. 0054, 2017.

[3] C. E. Madan, A. Kucukyilmaz, T. M. Sezgin, and C. Basdogan, "Recognition of haptic interaction patterns in dyadic joint object manipulation," *IEEE Transactions on Haptics*, vol. 8, pp. 54–66, 2015.

[4] 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 Confference of the IEEE*, vol. 3. IEEE, 2000, pp. 1773–1778.

[5] K. B. Reed and M. A. Peshkin, "Physical collaboration of human-human and human-robot teams," *IEEE Trans. Haptics*, vol. 1, no. 2, pp. 108–120, 2008.

[6] N. Stefanov, A. Peer, and M. Buss, "Role determination in human-human interaction," in *WHC'09: IEEE World Haptics Conference*, Salt Lake City, USA, 2009, pp. 51–56.

[7] A. Bussy, A. Kheddar, A. Crosnier, and F. Keith, "Human-humanoid haptic joint object transportation case study," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, 2012, pp. 3633–3638.

[8] E. Noohi, M. Žefran, and J. L. Patton, "A model for human–human collaborative object manipulation and its application to human–robot interaction," *IEEE transactions on robotics*, vol. 32, no. 4, pp. 880–896, 2016.

[9] E. A. Mielke, E. C. Townsend, and M. D. Killpack, "Analysis of rigid extended object co-manipulation by human dyads: Lateral movement characterization," *arXiv preprint arXiv:1702.00733*, 2017.

[10] A. Melendez-Calderon, V. Komisar, G. Ganesh, and E. Burdet, "Classification of strategies for disturbance attenuation in human-human collaborative tasks," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC*. IEEE, 2011, pp. 2364–2367.

[11] N. Jarrassé, T. Charalambous, and E. Burdet, "A framework to describe, analyze and generate interactive motor behaviors," *PLOS One*, vol. 7, no. 11, p. e49945, 2012.

[12] A. Kucukyilmaz, T. Sezgin, and C. Basdogan, "Intention recognition for dynamic role exchange in haptic collaboration," *IEEE Transactions on Haptics*, vol. 6, no. 1, pp. 58–68, 2013.

[13] A. Kucukyilmaz, S. O. Oguz, T. M. Sezgin, and C. Basdogan, "Improving human-computer cooperation through haptic role exchange and negotiation," in *Immersive Multimodal Interactive Presence*. Springer, 2012, pp. 229–254.

[14] 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.*, vol. 31, no. 13, pp. 1656–1674, 2012.

[15] J. R. Medina, T. Lorenz, and S. Hirche, "Synthesizing anticipatory haptic assistance considering human behavior uncertainty," *IEEE Transactions on Robotics*, vol. 31, no. 1, pp. 180–190, 2015.

[16] D. A. Abbink, M. Mulder, and E. R. Boer, "Haptic shared control: smoothly shifting control authority?" *Cognition, Technology & Work*, vol. 14, no. 1, pp. 19–28, 2012.

[17] Y. Li, K. P. Tee, W. L. Chan, R. Yan, Y. Chua, and D. K. Limbu, "Continuous role adaptation for human–robot shared control," *IEEE Transactions on Robotics*, vol. 31, no. 3, pp. 672–681, 2015.

[18] A. D. Dragan and S. S. Srinivasa, "A policy-blending formalism for shared control," *The International Journal of Robotics Research*, vol. 32, no. 7, pp. 790–805, 2013.

[19] M. B. Dias, B. Kannan, B. Browning, E. Jones, B. Argall, M. F. Dias, M. Zinck, M. M. Veloso, and A. Stentz, "Sliding autonomy for peer-to-peer human-robot teams," in *Proceedings of the international conference on intelligent autonomous systems*, 2008, pp. 332–341.

[20] D. Powell and M. K. O'Malley, "The task-dependent efficacy of shared-control haptic guidance paradigms," *IEEE transactions on haptics*, vol. 5, no. 3, pp. 208–219, 2012.

[21] S. O. Oguz, A. Kucukyilmaz, T. M. Sezgin, and C. Basdogan, "Supporting negotiation behavior with haptics-enabled human-computer interfaces," *IEEE transactions on haptics*, vol. 5, no. 3, pp. 274–284, 2012.

[22] H. C. Lin, I. Shafran, D. Yuh, and G. D. Hager, "Towards automatic skill evaluation: Detection and segmentation of robot-assisted surgical motions," *Computer Aided Surgery*, vol. 11, no. 5, pp. 220–230, 2006.

[23] N. Stefanov, A. Peer, and M. Buss, "Online intention recognition for computer-assisted teleoperation," in *2010 IEEE International Conference on Robotics and Automation*. IEEE, 2010, pp. 5334–5339.

[24] R. T. Olszewski, "Generalized feature extraction for structural pattern recognition in time-series data," Ph.D. dissertation, Carnegie-Mellon University, School of Computer Science, Pittsburgh, PA, US, 2001.

[25] M. Rank and M. Di Luca, "Speed/accuracy tradeoff in force perception." *Journal of experimental psychology: human perception and performance*, vol. 41, no. 3, p. 738, 2015.

[26] R. Groten, D. Feth, H. Goshy, A. Peer, D. Kenny, and M. Buss, "Experimental analysis of dominance in haptic collaboration," in *The 18th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) 2009*, 2009, pp. 723–729.

[27] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, ser. Adaptive computation and machine learning series. The MIT Press, 2006.