Communication Through Physical Interaction: A Study of Human Collaborative Manipulation of a Planar Object

Maria Javaid¹, Miloš Žefran¹ and Barbara Di Eugenio²

Abstract—In this paper we describe our progress towards understanding human communication through physical interaction. We describe a classification algorithm that can recognize four classes of actions that frequently occur during collaborative manipulation of planar objects. These actions were selected based on a user study involving dyads of elderly and care-giver in a realistic setting. Further user studies were conducted to collect the data necessary to develop the classification algorithm. As part of the data collection we also developed a sensory glove. The classification algorithm gives insight into human collaborative manipulation. More precisely, it identifies features in the data that are significant for classification. This information is particularly interesting as it only relies on physical aspects of the interaction and not on any particular sensor. As a result, the described work does not depend on any particular hardware and can be directly used by other researchers in human-robot interaction to develop further experiments and studies.

I. INTRODUCTION

Personal robot assistants hold great promise for addressing pressing societal needs. One of the areas where they could potentially have an enormous impact is to support the independent living of the elderly [1]–[3]. However, if a robot is to help an elderly person with activities of daily living (ADLs) [4], it needs to physically interact with the person. But physical interaction, be it through a direct touch or via an object, is also a form of communication that complements language, vision and gestures. This work is motivated by the need to investigate the communicative aspect of physical interaction so that future robots may be endowed with such a capability. While physical interaction between robots and humans has been well studied (e.g. [5]–[7]), the focus has been on interpreting the interaction at the control level rather than explore its communicative aspects.

In this work, we advance the hypothesis that physical interaction is an important communication modality that complements language, vision and gestures. For example, one can envision taking a hand of an elderly person with impaired vision and place it on a fork so the person can grab it, all without speaking a word. In [8], [9], we showed that physical interaction significantly improves the understanding of language.

Learning from humans is one of the prevailing paradigms in the human-robot interaction community [10]–[12]. Consistent with this view, in order to allow personal robots to communicate with humans through physical interaction, it is necessary to understand how humans communicate with each other in this way. This calls for human studies and data collection efforts. Our paper describes a set of human studies that were designed to establish a corpus of multimodal interactions between a subject and a helper, placing special emphasis on physical interaction. To start with, we conducted a user study involving dyads of elderly and care-giver in a realistic setting. The data from the study was used to identify activities where physical interaction plays an important role and focus additional data collection. Four classes of actions that frequently occur during collaborative manipulation of planar objects were chosen for further study. In turn, the bulk of the data used in this paper was collected in a laboratory setting, focusing on collaborative manipulation of planar objects and using younger subjects. A unique challenge in studying human physical interaction is the lack of devices that can collect data and are unobtrusive. As part of our data collection efforts we thus also developed a sensory glove for observing human physical interaction.

Our main result are decision rules that successfully identify the four classes of manipulative actions from the data. They were automatically learned using principal component analysis and clustering. An important feature of the decision rules is that they directly use physical aspects of the interaction and are not dependent on any particular hardware. The analysis thus provides a useful insight into human collaborative manipulation of planar objects. More importantly, it can be easily generalized to other devices and used by other researchers to develop further studies.

More broadly, our research is an important step towards developing personal robots that are engaged in both cognitive as well as physical aspects of the daily activities of the elderly. It is part of the RoboHelper project [13] whose ultimate goal is to deploy robotic assistants for the elderly so that they can safely remain living in their home.

II. RELATED WORK

In recent years, considerable effort has been devoted to improving the assistive technology for the elderly [14]–[16]. While some of these robots use touch sensors, the sensors have very limited function [17]–[19]. A number of studies have explored various aspects of haptic collaboration [20]–[26]. None of these works identifies mechanisms for communication through physical interaction in everyday activities. A major obstacle in studying physical interaction is a lack of suitable sensors. This may be attributed to both the complicated nature of the sense of touch (tactile and
kinesthetic) and the fact that the existing tactile sensors are limited. A variety of sensor designs have been proposed based on different transducers and materials [27]–[31]. The commercially available physical interaction sensing devices range from very precise force and torque sensors [32], [33] to flexible pressure sensors [34]–[36]. These give a rough estimate of the pressure at the point of contact rather than a precise measurement of force and torque. The big advantage of these sensors is that they are inexpensive and are suitable for use on sensory gloves. Among the data gloves currently available on the market, a particularly attractive choice is the Grip System [37]. It has been observed that the calibration of the sensors on this glove is fairly inaccurate [38]. Another option, Vista Medical Glove [39], is also not suitable for our study as the conducting strips attached to the sensors are too long and they might affect how the care-giver can perform activities. Another disadvantage of these gloves is that they only offer tactile sensing. Most notably, they lack inertial navigation sensors that could provide information about the orientation of the hand. This information is useful in distinguishing various grasp configurations.

III. NEWLY DEVELOPED SENSORY GLOVE

Given the lack of suitable devices on the market we decided to develop our own sensory glove. Since no suitable force/torque sensors are easily available, we used pressure sensors to obtain interaction force information. Two design requirements were identified for studying human physical interaction: (a) the sensing device should be able to measure direct or indirect (through an object) forces; and (b) it should not interfere with the observed activity.

Satisfying (a) would require placing sensors all over the surface of the arm. For example, in the user study described in [13], the caregiver uses her arm to help the elderly subject. Clearly, measuring all such interactions is impractical so we restricted ourselves to the hands, as that is where the majority of communication through physical interaction takes place. To further simplify the data acquisition process we only focused on the right hand. To satisfy (b) we used a wearable glove with pressure sensors. This requirement also calls for the necessary electronics to be lightweight and for the sensors on the glove to be flexible and thin so that the sensation of the subject and the movement of the hand are not significantly affected.

We chose a plain cotton glove, since it is comfortable to wear. To measure physical interaction data we attached iForce pressure sensors [40] (Tekscan, USA) to the glove. These sensors are thin and light, and favorably compare to other similar sensors in terms of precision and linearity [34], [35]. To minimize the interference of the sensors with the normal use of the hand we put the pressure sensitive part on the front side of the hand, wrapping the sensor to the backside of the hand where all the wires are attached. The sensors are stitched to the glove to hold them firmly in place.

The sensors were placed on every segment of each finger except for the middle segments of the thumb and the pinkie (these segments are too small). We also placed four of the pressure sensors on the palm. In total, 17 pressure sensors were attached to the glove (Figure 1). In addition to the pressure sensors, a 6 degree of freedom inertial measurement unit (ITG3200/ADXL345, SparkFun Electronics, USA [41]) was used to capture hand tilt and acceleration. It was attached to the back of the hand. The glove is connected to a processor box based on Arduino Mega microcontroller board [42] through two 20 wire cables. All the electronics is placed in a small backpack that the subject wears during the experiments.

Fig. 1: A sensory glove developed at UIC Robotics Lab

The microcontroller time stamps the data and transmits it wirelessly to a computer using an Xbee module. The data is sampled at around 70Hz.

The sensors cover most of the hand and do not hinder the bending of fingers. However, the pressure sensors are sensitive to pressure from both sides so pressure is recorded when fingers are bent. Further, the sensitivity of the pressure sensors varies among the sensors. We thus calibrated each sensor based on the maximum and minimum reading obtained during the experiments.

We should also mention that it is difficult to compare the performance of the developed glove to that of the commercial gloves due to a lack of evaluation studies. However, in Section VI-C we present a classification algorithm that was derived using the developed glove, but that does not depend on the particular hardware. This demonstrates that the developed glove is well suited for the intended use and further underscores the contribution of our work.

IV. USER STUDY

In order to better understand different communication modalities and types of interactions between the elderly and their caregivers, we conducted a user study as a part of the RoboHelper project [13]. The user study was performed in a fully functional studio apartment in the College of Nursing at Rush University. Our experiments focused mainly on the activities of the daily living (ADLs) that are crucial for the independent living of the elderly: (a) getting up from the bed/chair; (b) ambulating in the apartment; (c) cooking a meal; and (d) setting a table for a meal and subsequently cleaning up. During the experiments, video streams from 8
cameras were recorded to provide complementary views of the room and the subjects. The subjects also wore wireless microphones to record the audio. We collected 19 interactions, where each different elderly subject interacted with one of two helpers.

![Image of data collection](image)

**Fig. 2: Data collection in a mock apartment.**

To obtain the information on physical interaction, the helpers wore the data glove equipped with pressure sensors described in Section III on their right hand. While the data glove only provides limited information on the forces during the physical interaction, it only minimally interferes with the normal interaction between subjects. None of the subjects ever complained that they could not perform an ADL properly because they wore the data glove.

The experiments confirmed our hypothesis that physical interaction plays an important role in the communication of the elderly with the caregiver as reported in [13]. The activities that require physical interaction can be divided into the following broad categories: (a) handing-over objects; (b) manipulating an object together; and (c) supporting the elderly in walking or getting up.

It was also observed that both handing-over objects and manipulating objects together often involved planar objects, e.g. handing over plates or a tray while setting and cleaning the table. This was the case for 15 subjects out of 19 subject. We thus decided to further investigate collaborative manipulation of planar objects. Our preliminary data suggests that collaborative manipulation of planar objects mainly consists of the following actions:

1) Holding an object alone with one hand (OH).
2) Holding an object alone with both hands (BH).
3) Holding an object with another person (AP).
4) Empty hand—not holding anything (EH).

In order to advance our understanding of multimodal communication it is necessary to establish an annotated corpus of such interactions. For example, it has been shown that knowing the type of physical interaction helps in understanding spoken language [8], [9]. Since manual annotation is time consuming and does not scale, it is thus necessary to develop appropriate classification algorithms that can be used to automatically annotate the data. In turn, if implemented on a robot, the classification algorithms can be used during the interaction to guide robot actions.

V. DATA COLLECTION IN A LABORATORY SETTING

Since our preliminary data collection was largely unscripted and took place in an unstructured setting, we decided to collect additional data in a laboratory setting. The experiments consisted of one person wearing the data glove and performing different instances of the planar manipulation task in collaboration with another person in whatever order they preferred. The object that was manipulated was a dinner plate. For example, the subject wearing the glove held the plate alone for some time with the gloved hand, then for some time held it with the collaborating subject, and so on. We performed each of the four actions mentioned earlier in a completely random order. Since these actions were random, certain actions were performed for a longer duration than others, and some actions were repeated more than others, but each action was repeated at least ten times. To filter out the noise, the recorded data was filtered using a moving average smoothing filter. The filtered data is subsequently down-sampled to 20Hz. For each action, each sample is considered a separate data point. For example, if an instance of a particular action lasts for 3 seconds, we have 20×3 = 60 data points for that instance. In our classification experiments we classify each of these data points individually.

![Diagram of data processing](diagram)

**Fig. 3: Steps in processing of the collected data.**

The experimental data was collected from 4 subjects in total. Experiments were videotaped. The video was then synchronized with the glove data via the latter’s time stamps. Each sample was annotated for actions listed above based on the video. These annotations were used for verification of predicted results. Figure 3 describes the steps involved in the data processing.

VI. CLASSIFICATION OF DATA

Several classification algorithms were compared. The primary measures used for comparison were precision (the portion of instances that were assigned to a particular class that were correctly classified) and recall (the portion of all instances belonging to a class which were assigned to that class) [43]. The abbreviations used in the tables are those listed in Section IV.

A. Supervised Classification

We first classified the data using k-Nearest Neighbor (k-NN [44]) and Linear Discriminant Analysis (LDA) [45], two examples of supervised classification algorithms. Supervised classification algorithms require training data, where classes for each sample have been labeled, to classify the testing data. In our experiment, we used the data from 3 subjects as training data and the data from the remaining subject as testing data. This was repeated for each of the four
subjects. The data was classified into the four classes listed in Section IV (OH, BH, AP, EH). The average F-score (the harmonic mean of precision and recall) were 52.15% and 56.93% for LDA and k-NN, respectively.

B. Unsupervised Classification

We next explored clustering, a representative of unsupervised classification algorithms. Clusters are formed so that the objects in the same cluster are more closely related to each other than to those in other clusters [44]. When we applied clustering directly to the 17 dimensional data comprised of pressure sensor readings the results were not good. We thus reduced the dimensionality of the data by applying principal component analysis (PCA) [46].

The results of PCA show that more than 50% of the total information is concentrated in the first two principal components [45]. Figure 4 shows the data of subject 3 projected to two leading principal components and indicates that the data is clustered.

![Data plot of two leading principal components for subject 3](image)

Fig. 4: Data plot of two leading principal components for subject 3

We chose the k-means clustering algorithm [44] and applied it to the data represented by the first two principal components. To select the number of clusters we used mean silhouette value [47] which determines how well the data points have been clustered. We applied clustering to the data of subject 3. In this case, the mean silhouette value was highest for 12 clusters. The results are given in Table I.

**TABLE I: Data Composition of 12 Clusters for Subject 3**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>AP</td>
<td>2</td>
<td>571</td>
<td>68</td>
<td>142</td>
<td>289</td>
<td>234</td>
<td>29</td>
<td>241</td>
<td>0</td>
<td>429</td>
<td>2</td>
<td>526</td>
</tr>
<tr>
<td>OH</td>
<td>2</td>
<td>279</td>
<td>0</td>
<td>650</td>
<td>981</td>
<td>145</td>
<td>84</td>
<td>18</td>
<td>0</td>
<td>517</td>
<td>699</td>
<td>197</td>
</tr>
<tr>
<td>EH</td>
<td>0</td>
<td>21</td>
<td>251</td>
<td>0</td>
<td>4</td>
<td>808</td>
<td>330</td>
<td>145</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BH</td>
<td>458</td>
<td>83</td>
<td>12</td>
<td>1</td>
<td>1094</td>
<td>573</td>
<td>130</td>
<td>171</td>
<td>0</td>
<td>31</td>
<td>0</td>
<td>445</td>
</tr>
</tbody>
</table>

Note that the clusters 1-6, 9 and 11 contain data samples predominantly from a single action whereas the clusters 7, 8, 10 and 12 confuse two or more actions. Recall that the numbers in Table I represent individual data samples; the numbers are high since there are 20 samples for each second of the experiment.

Table II gives the precision and recall for the well separated clusters for Subject 3 and also the average for all the subjects. Recall is low, as we have not considered clusters that are shared by two or more actions.

**TABLE II: Clustering Results for Subject 3 and Average Across All Subjects**

<table>
<thead>
<tr>
<th>Subject 3</th>
<th>All Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>Recall</td>
</tr>
<tr>
<td>AP</td>
<td>80.0%</td>
</tr>
<tr>
<td>OH</td>
<td>59.9%</td>
</tr>
<tr>
<td>EH</td>
<td>83.9%</td>
</tr>
<tr>
<td>BH</td>
<td>96.0%</td>
</tr>
<tr>
<td></td>
<td>77.0%</td>
</tr>
<tr>
<td></td>
<td>76.7%</td>
</tr>
<tr>
<td></td>
<td>90.0%</td>
</tr>
<tr>
<td></td>
<td>76.1%</td>
</tr>
</tbody>
</table>

C. Clustering Based Decision Tree (CBDT)

Clustering using PCA gives satisfactory results across subjects, but it unfortunately provides little insight into the physical features that characterize different classes. Even more troublesome, the results are hardware-specific and cannot be generalized to a different data-collection device.

To obtain a more meaningful interpretation of the clusters, we thus mapped the first principal component to the sensors that define it. Namely, we identified the sensors that are given more weight in the leading principal component. This component carries 30%-60% of total information for the data in the 4 experiments. We determined that the highest weight in the leading principal component is given to fingertip sensors, excluding the index finger, for all the subjects. By interpreting the weights that define well separated clusters, we subsequently built a decision tree shown in Figure 5. Please refer to Table III for explanation of different pressure levels. Note that the analog inputs on the Arduino microcontroller are sampled using 10-bit A/D converters which implies that the maximum sensor reading is 1023.

**TABLE III: Pressure Level Definitions**

<table>
<thead>
<tr>
<th>Sum of pinkie, ring and middle finger top</th>
<th>Very High</th>
<th>High</th>
<th>Moderate</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>greater than 2100</td>
<td>greater than 1800 and greater than 700</td>
<td>less than 1800</td>
<td>less than 700</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pinkie Top</th>
<th>Very High</th>
<th>High</th>
<th>Very Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA</td>
<td>NA</td>
<td>less than 150</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ring Top</th>
<th>Very High</th>
<th>High</th>
<th>Very Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>greater than 900</td>
<td>NA</td>
<td>less than 300</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Middle Top</th>
<th>Very High</th>
<th>High</th>
<th>Very Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>greater than 900</td>
<td>greater than 700</td>
<td>less than 300</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Thumb Top</th>
<th>Very High</th>
<th>High</th>
<th>Very Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>greater than 750</td>
<td>NA</td>
<td>less than 300</td>
<td></td>
</tr>
</tbody>
</table>

The physical interpretation of these rules is apparent. For example, a point will be classified as empty hand either if the pressure on the fingertips is very low (the fingertips are not touching anything), or if the pressure is low but the palm is bent. The latter can be determined by checking the palm middle sensor. When the hand is empty it is often slightly closed, causing the palm middle sensor to bend, and in turn producing high pressure on that sensor.

Similarly, if there is some pressure on the tips of the three middle fingers, but on one of these fingers the pressure is
very low, the sample likely corresponds to empty hand if the palm is bent, and to holding a planar object (a plate) alone with both hands otherwise. In the latter case, when a person is holding a planar object such as a plate with both hands, one often relaxes the pressure on one or two of the fingers as one knows that the other hand is also holding the plate. The sensor readings that correspond to the other two classes can be similarly explained.

For the CBDT, Table IV provides performance and the confusion matrix. In the confusion matrix, each column represents the number of samples that were classified in the class described by the column label, while each row indicates how samples in the class described by the row label were classified. The average F-score (the harmonic mean of precision and recall) for the classification with the CBDT is 69.94%. We thus conclude that the CBDT successfully distinguishes different actions during human collaborative manipulation. It is especially interesting that using the pressure data it is possible to distinguish whether an object is held with two hands by a single person, or by two different people.

Table IV: Recognition Results for CBDT

<table>
<thead>
<tr>
<th>Recognition Rates</th>
<th>Confusion Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision Recall</td>
<td>AP</td>
</tr>
<tr>
<td>AP 57.2% 68.0%</td>
<td>11064</td>
</tr>
<tr>
<td>OH 71.9% 54.2%</td>
<td>6599</td>
</tr>
<tr>
<td>EH 88.5% 88.9%</td>
<td>163</td>
</tr>
<tr>
<td>BH 63.7% 70.8%</td>
<td>1302</td>
</tr>
</tbody>
</table>

The results of the CBDT across all subjects compare favorably with the results of \(k\)-NN for a single subject. This implies that the inferred physical interpretation of the data and the derived decision tree captures the information remarkably well.

VII. CONCLUSIONS

This paper investigates how humans communicate through physical interaction. The work is motivated by the need to understand human behavior before a similar functionality can be replicated on robots. The task that was studied in detail was collaborative manipulation of a planar object. We describe a sensory glove that was developed to unobtrusively capture features of the interaction forces during collaborative manipulation. We also describe a set of experiments with human subjects that show that physical interaction does in fact represent a form of communication. The ensuing data analysis allowed us to derive a decision tree that uses the rules which only depend on direct physical interpretation of the data (fingertip pressure). Different actions which seem alike looking at the forces required for manipulation of the object (holding with two hands by a single person rather than holding by two people) can be successfully recognized as distinct events using the derived decision tree. Since the rules in the decision tree only use relative pressure to distinguish between different actions, they can be easily adapted for different sensors and hardware platforms. In fact, one would expect that the recognition rates would only improve if better sensing hardware is used.

The findings in the paper can be directly used to improve the ability of the robots to physically interact with humans. For example, the classes that were identified in this work can be used by the robot to determine how to act during a hand-over task with a human. Implementation on a robotic platform is part of our future work.

While the actions studied in this paper all involved power planar grasp we believe that the derived decision tree could be generalized to actions that require other types of grasps due to the direct physical interpretation of the decision rules. Showing that this is indeed the case is another promising extension of this work.
REFERENCES


